

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 23, 2005, 12:25:43 ; Search time 121 Seconds
(without alignment)
1168.048 Million cell updates/sec

Title: US-10-736-892-12

Perfect score: 276

Sequence: 1 MAAGFKTVEPLEYYRFLKE.....TRHKVSKLLDEVITQSMKHK 276

Scoring table:  Gapop 60.0 , Gapext 60.0

Searched: 1612378 seqs, 512079187 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1612378

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : UniProt_03.*

1: uniprot_sprot.*

2: uniprot_trembl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	78	28.3	270	Q6NX62	Q6NX62 mus musculus
2	78	28.3	276	1 R443 MOUSE	Q94753 mus musculus
3	78	28.3	276	1 Q6PCW8	Q6PCW8 mus musculus
4	69	25.0	276	1 R443 HUMAN	Q96b26 homo sapien
5	20	7.2	276	2 Q8AVT6	Q8AVT6 xenopus lae
6	19	6.9	276	2 Q6KK0	Q6KK0 xenopus tro
7	17	6.2	277	2 Q6DRN4	Q6DRN4 brachydanio
8	10	3.6	275	1 ECX2 SULSO	Q9uxc0 sulfolobus
9	10	3.6	277	1 ECX2 PYRFU	Q8u0m0 pyrococcus
10	8	2.9	259	1 ECX2 ARCFU	Q29756 archaeoglob
11	8	2.9	274	1 ECX2 PYRAB	Q9v118 pyrococcus
12	8	2.9	274	1 ECX2 PYRHO	Q59224 pyrococcus
13	8	2.9	275	1 ECX2 SULTO	Q975g9 sulfolobus
14	8	2.9	276	1 ECX2 AERPE	Q9yc05 aeropyrum p
15	8	2.9	280	2 Q700D9	Q7Q0d9 anopheles g
16	8	2.9	302	2 Q9ZUI4	Q9ZUI4 arabidopsis
17	8	2.9	308	2 Q93QD4	Q93qd4 staphylococ
18	8	2.9	308	2 Q99UN8	Q99un8 staphylococ
19	8	2.9	308	2 Q7A124	Q7A124 staphylococ
20	8	2.9	308	2 Q7A5Z3	Q7A5z3 staphylococ
21	8	2.9	308	2 Q6G9Y3	Q6G9y3 staphylococ
22	8	2.9	308	2 Q6GKH5	Q6Gkh5 staphylococ
23	8	2.9	336	2 Q67FU6	Q67pu6 symbiobacte
24	8	2.9	347	2 Q6BHU9	Q6bhu9 debaryomyce
25	8	2.9	415	2 Q7V3D5	Q7v3d5 prochloroco
26	8	2.9	990	2 O13936	O13936 schizosacch
27	8	2.9	1666	2 Q7R3J9	Q7r3j9 giardia lam
28	8	2.9	5017	2 Q63DF3	Q63df3 bacillus ce
29	8	2.9	5017	2 Q6HKW5	Q6hkw5 bacillus th
30	7	2.5	76	2 Q6Q155	Q6q155 rattus norv
31	7	2.5	82	2 Q9H396	Q9h396 homo sapien

32	7	2.5	83	2	O26727	O26727 methanobact
33	7	2.5	86	2	Q84622	Q84622 paramecium
34	7	2.5	88	2	Q6KFT9	Q6kft9 ateles geof
35	7	2.5	88	2	Q8KKD2	Q8kkd2 helicobacte
36	7	2.5	88	2	Q8KKE2	Q8kke2 helicobacte
37	7	2.5	94	2	Q7LIR6	Q7lir6 lactobacill
38	7	2.5	96	2	Q8R401	Q8r401 rattus norv
39	7	2.5	105	2	Q9AJG8	Q9ajg8 vibrio prot
40	7	2.5	105	2	O06156	O06156 mycobacteri
41	7	2.5	105	2	Q7MNV8	Q7mnv8 vibrio vuln
42	7	2.5	105	2	Q7TW46	Q7tw46 mycobacteri
43	7	2.5	105	2	Q87SH1	Q87sh1 vibrio para
44	7	2.5	105	2	Q8DEK3	Q8dek3 vibrio vuln
45	7	2.5	106	2	Q743X4	Q743x4 mycobacteri

ALIGNMENTS

RESULT 1
Q6NX62 PRELIMINARY; PRT; 270 AA.
AC Q6NX62;
DT 05-JUL-2004 (TREMBlrel. 27, Created)
DT 05-JUL-2004 (TREMBlrel. 27, Last sequence update)
DT 05-JUL-2004 (TREMBlrel. 27, Last annotation update)
DE Exosc8 protein (Fragment).
GN Name=Exosc8;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
NCBI_TaxID=10090;
[1]
RN SEQUENCE FROM N.A.
RC TISSUE=Mammary gland;
RX MEDLINE=2238625; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strauberg R.L., Feingold E.A., Grouse L.H., Berge J.G.,
Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
Brownstein M.J., Ustin T.B., Toshiyuki S., Carninci P., Prange C.,
Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
Fahey J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,
Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
Krzywinski M.I., Skalska U., Smailus D.E., Schnerch A., Schein J.E.,
Jones S.J., Marra M.A.
RT "Generation and initial analysis of more than 15,000 full-length human
and mouse cDNA sequences.";
Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
[2]
RN SEQUENCE FROM N.A.
RC TISSUE=Mammary gland;
RA Strauberg R.;
RL Submitted (MAR-2004) to the ENBL/GenBank/DBJ databases.
EMBL; BC067250; AAH67250.1; -.
DR GO; GO:000175; F:3'-5'-exoribonuclease activity; IEA.
DR GO; GO:0003723; F:RNA binding; IEA.
DR GO; GO:0006396; F:RNA processing; IEA.
DR InterPro; IPR001247; 3_ExonRase.
DR Pfam; PF01138; RNase PH; 1.
DR Pfam; PF03725; RNase PH_C; 1.
FT NON_TER
SQ SEQUENCE 270 AA; 29343 MW; 338EB40913D55696 CRC64;
Query Match 28.3%; Score 78; DB 2; Length 270;
Best Local Similarity 100.0%; Pred. No. 1.9e-69;

Matches 78; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 35 TTVNIGSISTADGSAALVKGNTTTCGKAEFAAPPVDPDRGYVVPVNDLPPCSSLRFR 94
|||||
Db 29 TTVNIGSISTADGSAALVKGNTTTCGKAEFAAPPVDPDRGYVVPVNDLPPCSSLRFR 88
|||||

QY 95 TGPGEAQTTSQFIADV 112
|||||
Db 89 TGPGEAQTTSQFIADV 106
|||||

RESULT 2

RR43 MOUSE

ID RR43 MOUSE STANDARD; PRT; 276 AA.

AC Q9D753;

DT 28-FEB-2003 (Rel. 41, Created)

DT 28-FEB-2003 (Rel. 41, Last sequence update)

DT 05-JUL-2004 (Rel. 44, Last annotation update)

DE Exosome complex exonuclease RRP43 (EC 3.1.13.-) (Ribosomal RNA processing protein 43) (Exosome component 8).

DE Processing protein 43 (Exosome component 8).

GN Name=Exosc8; Synonym=Rrp43;

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OX NCBI_TaxID=10090;

RN [1]

SEQUENCE FROM N.A.

RP STRAIN=C57BL/6J; TISSUE=Tongue;

RC MEDLINE=22354683; PubMed=12466851; DOI=10.1038/nature01266;

RX Okazaki Y., Furuno M., Kasukawa T., Adachi J., Bono H., Kondo S., Naito I., Oatso N., Saito R., Suzuki H., Yamanaka I., Kiyosawa H., Yagi K., Tomaru Y., Hasegawa Y., Nogami A., Schonbach C., Gojobori T., Baldarelli R., Hill D.P., Bult C., Hume D.A., Quackenbush J., Schriml L.M., Kanapin A., Matsuda H., Batalov S., Beisel K.W., Blake J.A., Bradt D., Brusic V., Chothia C., Corbani L.E., Cousins S., Dalla E., Dragani T.A., Fletcher C.F., Forrest A., Frazer K.S., Gaasterland T., Gariboldi M., Gissi C., Godzik A., Gough J., Grimmond S., Gustincich S., Hirokawa N., Jackson I.J., Jarvis E.D., Kanai A., Kawaji H., Kawasawa Y., Kedzierski R.M., King B.L., Konggaya A., Kurochkin I.V., Lee Y., Lenhard B., Lyons P.A., Maglott D.R., Maltais L., Marchionni L., McKenzie L., Miki H., Nagashima T., Numata K., Okido T., Pavan W.J., Perle G., Pesole G., Petrovsky N., Pillai R., Pontius J.U., Qi D., Ramachandran S., Ravasi T., Reed J.C., Reed D.J., Reid J., Ring B.Z., Ringwald M., Sandelin A., Schneider C., Sempie C.A., Setou M., Shimada K., Sultana R., Takenaka Y., Taylor M.S., Teasdale R.D., Tomita M., Verardo R., Wagner L., Wahlestedt C., Wang Y., Watanabe Y., Wells C., Wilming L.G., Wyshaw-Boris A., Yanagisawa M., Yang I., Yang L., Yuan Z., Zavolan M., Zhu Y., Zimmer A., Carninci P., Hayatsu N., Hirozane-Kishikawa T., Konno H., Nakamura M., Sakazume N., Sato K., Shiraki T., Waki K., Kawai J., Aizawa K., Arakawa T., Fukuda S., Hara A., Hashizume W., Imotani K., Ishii Y., Itoh M., Kagawa I., Miyazaki A., Sakai K., Sasaki D., Shibata K., Shinagawa A., Yasunishi A., Yoshino M., Waterston R., Lander E.S., Rogers J., Birney E., Hayashizaki Y.

RT "Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs."

RT Nature 420:563-573 (2002).

CC -!- FUNCTION: Component of the nuclear exosome 3'-5' exoribonuclease complex. Required for the 3' processing of the 7S pre-RNA to the mature 5.8S rRNA. Has a 3'-5' exonuclease activity (By similarity).

CC -!- SUBUNIT: Component of the exosome multienzyme ribonuclease complex composed of at least 11 proteins: RRP4, RRP40, RRP41/SK16, RRP42, RRP43, RRP44/DIS3, PM/SCI-75, RRP46, CSL4 and PM/SCI-100 (only in the nuclear complex). Also associated with the GTPase Ran (By similarity).

CC -!- SUBCELLULAR LOCATION: Cytoplasmic and nuclear; nucleolar (By similarity).

CC -!- SIMILARITY: Belongs to the RNase PH family.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC -----

DR EMBL; AK009584; -; NOT ANNOTATED_CDS.

DR MGD; MGI:1916889; 2310032N20R1X.

DR InterPro; IPR001247; 3_ExoRNase.

DR Pfam; PF01138; RNase_PH_1.

DR Pfam; PF03725; RNase_PH_C_1.

DR Exonuclease; Exosome; Hydrolase; Nuclear protein; Nuclease;

KW RNA-binding; rRNA processing.

SQ SEQUENCE 276 AA; 29949 MW; 297E1E45F5C0F794 CRC64;

Query Match 28.3%; Score 78; DB 1; Length 276;

Best Local Similarity 100.0%; Pred. No. 1.9e-69;

Matches 78; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 35 TTVNIGSISTADGSAALVKGNTTTCGKAEFAAPPVDPDRGYVVPVNDLPPCSSLRFR 94
|||||

Db 35 TTVNIGSISTADGSAALVKGNTTTCGKAEFAAPPVDPDRGYVVPVNDLPPCSSLRFR 94
|||||

QY 95 TGPGEAQTTSQFIADV 112
|||||

Db 95 TGPGEAQTTSQFIADV 112
|||||

RESULT 3

Q6PCW8

ID Q6PCW8 PRELIMINARY; PRT; 276 AA.

AC Q6PCW8;

DT 05-JUL-2004 (TRENBLrel. 27, Created)

DT 05-JUL-2004 (TRENBLrel. 27, Last sequence update)

DT 05-JUL-2004 (TRENBLrel. 27, Last annotation update)

DE CBP-interacting protein 3.

DE Name=Exosc8;

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OX NCBI_TaxID=10090;

RN [1]

SEQUENCE FROM N.A.

RP TISSUE=Limb.

RC MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.2426038999;

RX Strausberg R.L., Feingold E.A., Grouse L.H., Dege J.G., Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D., Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K., Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F., Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L., Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E., Brownstein M.J., Udwin T.B., Toshiyuki S., Carninci P., Prange C., Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J., Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H., Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W., Villalón D.K., Wuzny D.M., Sodergren E.J., Lu X., Gibbs R.A., Fahey J., Helton E., Kettelman M., Madan A.C., Rodriguez S., Sanchez A., Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G., Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C., Rodriguez A.C., Grinwood J., Schmutz J., Myers R.M., Butterfield Y.S., Krzywicki M.I., Skalska U., Smailus D.E., Schnerch A., Schein J.E., Jones S.J., Marra M.A.

RT "Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences."

RT Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).

RN [2]

SEQUENCE FROM N.A.

RP TISSUE=Limb.

RC Strausberg R.

RL Submitted (OCT-2003) to the EMBL/GenBank/DBJ databases.

DR EMBL; BC059089; AAH59089.1; -

DR GO; GO:000175; F:3'-5'-exoribonuclease activity; IEA.

DR GO; GO:0003723; F:RNA binding; IEA.

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 23, 2005, 12:32:04 ; Search time 44 Seconds
(without alignments)
603.542 Million cell updates/sec

Title: US-10-736-892-12

Perfect score: 276

Sequence: 1 MAAGFKTVEPLEYYRFLKE.....TRHKVSKLDEVQSMKHK 276

Scoring table:  Gapop 60.0 , Gapext 60.0

Searched: 283416 seqs, 96216763 residues

Word size : 0

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : PIR 79:*

1: pir1.*

2: pir2.*

3: pir3.*

4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	10	3.6	275	G90221	conserved hypothet
2	8	2.9	259	F69311	conserved hypothet
3	8	2.9	274	F75181	polyribonucleotide
4	8	2.9	274	D71032	probable autoantig
5	8	2.9	276	E72623	probable autoantig
6	8	2.9	302	D96625	hypothetical prote
7	8	2.9	308	A89896	malonyl CoA-acyl c
8	8	2.9	990	T38274	probable transcrip
9	7	2.5	83	H69183	hypothetical prote
10	7	2.5	86	T17803	hypothetical prote
11	7	2.5	105	E70552	hypothetical prote
12	7	2.5	106	D82082	cell division prot
13	7	2.5	128	A83451	hypothetical prote
14	7	2.5	129	A86422	hypothetical prote
15	7	2.5	144	H70063	hypothetical prote
16	7	2.5	191	S56012	XS-2 protein (homo
17	7	2.5	193	D82385	mammotol operon re
18	7	2.5	218	C83152	hypothetical prote
19	7	2.5	222	B90096	hypothetical prote
20	7	2.5	227	T32894	hypothetical prote
21	7	2.5	244	T54911	hypothetical prote
22	7	2.5	246	G72623	probable ribonucle
23	7	2.5	262	D70434	tryptophan synthas
24	7	2.5	277	E84478	hypothetical prote
25	7	2.5	283	C59744	conserved hypothet
26	7	2.5	290	F90352	hypothetical prote
27	7	2.5	290	A93971	hypothetical prote
28	7	2.5	292	F69372	osmoprotection pro
29	7	2.5	317	H81245	glycerate dehydrog

ALIGNMENTS

RESULT 1

G90221 conserved hypothetical protein [imported] - Sulfolobus solfataricus
C:Species: Sulfolobus solfataricus

C>Date: 24-May-2001 #sequence_revision 24-May-2001 #text_change 09-Jul-2004

C:Accession: G90221

R:She, Q.; Singh, R.K.; Confalonieri, F.; Zivanovic, V.; Allard, G.; Awayez, M.J.; Chan-
Jong, I.; Jeffries, A.C.; Kozera, C.W.; Medina, N.; Peng, X.; Thi-Ngoc, H.P.; Redder, P.;
arrett, R.A.; Ragan, M.A.; Sensen, C.J.; Van der Oost, J.
submitted to GenBank, April 2001

A:Description: Sulfolobus solfataricus complete genome.

A:Reference number: A99139

A:Accession: G90221

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-275 <KUR>

A:Cross-references: UNIPROT:Q9UKC0; GB:AE006641; NID:gi3813902; PIDN:AAK41030.1; GSPDB:G
C:Genetics:

A:Gene: SS00732

C:Superfamily: conserved hypothetical protein MTH682

Query Match 3.6%; Score 10; DB 2; Length 275;
Best Local Similarity 100.0%; Pred. No. 0.069;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 45 ADGSALVKLG 54

|||||

Db 50 ADGSALVKLG 59

RESULT 2

F69311 conserved hypothetical protein AF0494 - Archaeoglobus fulgidus

C:Species: Archaeoglobus fulgidus

C>Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 09-Jul-2004

C:Accession: F69311

R:Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dodson
.; Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirkness, E.F.
Glodek, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L.

Nature 390, 364-370, 1997

A:Authors: Uterback, T.; Cotton, M.D.; Spriggs, T.; Attiach, P.; Kaine, B.P.; Sykes, S.
Smith, H.O.; Woese, C.R.; Venter, J.C.

A:Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing archaeo

A:Reference number: A69250; MUID:98049343; PMID:9389475

A:Accession: F69311

A>Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-259 <KLE>

A:Cross-references: UNIPROT:O29756; GB:AE001070; GB:AE000782; NID:g2689393; PIDN:AAB9074
C:Superfamily: conserved hypothetical protein MTH682

Query Match 2.9%; Score 8; DB 2; Length 259;

[illegible]

RESULT 6

C;Species: *Arabislopsis thaliana* (mouse-ear cress)
C;Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 09-Jul-2004
C;Accession: D96625

Chen, C.W.; Chung, M.K.;
ansen, N.F.; Hughes, B.;

A:Authors: Salzberg, S.L.; Schwartz, J.R.K.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon, R.; Zuo, M.; Rooney, I.; Rowley, D.; Sakano, H.
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: A86141; MID:21016719; PMID:11130712
A:Accession: D96625
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-302 <STO>
A:Cross-references: UNIPROT:Q9ZUI4; GB:AE005173; NID:G4249388; PIDN:AAD14485.1; GSPDB:GN C:Genetics:
A:Gene: T2K10.14
A:Map position: 1
C:Superfamily: conserved hypothetical protein MTH682

DD
37 STADGARD 64

malonyl CoA-acyl carrier]

A, Title: Whole genome sequencing of methicillin-resistant *Staphylococcus aureus*.
A, Reference number: A89758; PMID:21311952; PMID:1148146
A, Accession: A89896
C, Title: Whole genome sequencing of methicillin-resistant *Staphylococcus aureus*.
C, Reference number: A89758; PMID:21311952; PMID:1148146
C, Accession: A89896
L, Title: Whole genome sequencing of methicillin-resistant *Staphylococcus aureus*.
L, Reference number: A89758; PMID:21311952; PMID:1148146
L, Accession: A89896
M, Title: Whole genome sequencing of methicillin-resistant *Staphylococcus aureus*.
M, Reference number: A89758; PMID:21311952; PMID:1148146
M, Accession: A89896
P, Title: Whole genome sequencing of methicillin-resistant *Staphylococcus aureus*.
P, Reference number: A89758; PMID:21311952; PMID:1148146
P, Accession: A89896
S, Title: Whole genome sequencing of methicillin-resistant *Staphylococcus aureus*.
S, Reference number: A89758; PMID:21311952; PMID:1148146
S, Accession: A89896
T, Title: Whole genome sequencing of methicillin-resistant *Staphylococcus aureus*.
T, Reference number: A89758; PMID:21311952; PMID:1148146
T, Accession: A89896
V, Title: Whole genome sequencing of methicillin-resistant *Staphylococcus aureus*.
V, Reference number: A89758; PMID:21311952; PMID:1148146
V, Accession: A89896
W, Title: Whole genome sequencing of methicillin-resistant *Staphylococcus aureus*.
W, Reference number: A89758; PMID:21311952; PMID:1148146
W, Accession: A89896
X, Title: Whole genome sequencing of methicillin-resistant *Staphylococcus aureus*.
X, Reference number: A89758; PMID:21311952; PMID:1148146
X, Accession: A89896
Y, Title: Whole genome sequencing of methicillin-resistant *Staphylococcus aureus*.
Y, Reference number: A89758; PMID:21311952; PMID:1148146
Y, Accession: A89896
Z, Title: Whole genome sequencing of methicillin-resistant *Staphylococcus aureus*.
Z, Reference number: A89758; PMID:21311952; PMID:1148146
Z, Accession: A89896

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 23, 2005, 12:39:24 ; Search time 97 Seconds
(without alignments)
946.911 Million cell updates/sec

Title: US-10-736-892-12
Perfect score: 276
Sequence: 1 MAAGKTVPELYRRFLKE.....TRHKEVSKLDEVIQSMKHK 276

Scoring table: Gapop 60.0 , Gapext 60.0

Searched: 1424015 seqs, 332791073 residues

Word size : 0
Total number of hits satisfying chosen parameters: 1424015

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

- Database : Published Applications AA:*
- 1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
 - 2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep.*
 - 3: /cgn2_6/ptodata/1/pubpaa/US05_NEW_PUB.pep.*
 - 4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
 - 5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
 - 6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep.*
 - 7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
 - 8: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep.*
 - 9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep.*
 - 10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep.*
 - 11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep.*
 - 12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep.*
 - 13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep.*
 - 14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep.*
 - 15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep.*
 - 16: /cgn2_6/ptodata/1/pubpaa/US10D_PUBCOMB.pep.*
 - 17: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
 - 18: /cgn2_6/ptodata/1/pubpaa/US11_NEW_PUB.pep.*
 - 19: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
 - 20: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	46	16.7	108	15	US-10-108-260A-3952
2	8	2.9	302	9	US-09-815-242-5460
3	8	2.9	308	14	US-10-138-701-16
4	8	2.9	308	15	US-10-282-122A-43948
5	8	2.9	308	17	US-10-857-625-624
6	8	2.9	311	9	US-09-815-242-12130
7	7	2.5	34	16	US-10-437-963-126146
8	7	2.5	44	15	US-10-424-599-157053
9	7	2.5	49	15	US-10-424-599-249020
10	7	2.5	64	15	US-10-424-599-222295
11	7	2.5	76	15	US-10-282-122A-61691
12	7	2.5	80	15	US-10-424-599-246330
13	7	2.5	82	16	US-10-755-889-474

14	7	2.5	100	15	US-10-424-599-257213	Sequence 257213, A
15	7	2.5	105	15	US-10-282-122A-62284	Sequence 62284, A
16	7	2.5	105	15	US-10-282-122A-64904	Sequence 64904, A
17	7	2.5	109	15	US-10-433-256-11	Sequence 11, Appl
18	7	2.5	116	17	US-10-488-197-22	Sequence 22, Appl
19	7	2.5	124	16	US-10-767-701-54901	Sequence 54901, A
20	7	2.5	135	17	US-10-499-805-4	Sequence 4, Appl
21	7	2.5	140	14	US-10-141-645-66	Sequence 66, Appl
22	7	2.5	141	16	US-10-767-701-39555	Sequence 39555, A
23	7	2.5	142	15	US-10-425-114-70041	Sequence 70041, A
24	7	2.5	161	15	US-10-424-599-246084	Sequence 246084, A
25	7	2.5	162	15	US-10-424-599-253574	Sequence 253574, A
26	7	2.5	171	15	US-10-424-599-266687	Sequence 266687, A
27	7	2.5	181	15	US-10-424-599-246331	Sequence 246331, A
28	7	2.5	193	15	US-10-282-122A-77693	Sequence 77693, A
29	7	2.5	218	9	US-09-815-242-11644	Sequence 11644, A
30	7	2.5	223	15	US-10-424-599-166100	Sequence 166100, A
31	7	2.5	254	15	US-10-282-122A-59409	Sequence 59409, A
32	7	2.5	265	15	US-10-425-114-55950	Sequence 55950, A
33	7	2.5	275	15	US-10-424-599-284027	Sequence 284027, A
34	7	2.5	279	17	US-10-722-045-43	Sequence 43, Appl
35	7	2.5	285	16	US-10-437-963-189797	Sequence 189797, A
36	7	2.5	286	16	US-10-781-014-268	Sequence 268, App
37	7	2.5	307	9	US-09-738-626-3909	Sequence 3909, App
38	7	2.5	308	16	US-10-437-963-138106	Sequence 138106, A
39	7	2.5	317	15	US-10-282-122A-48025	Sequence 48025, A
40	7	2.5	321	9	US-09-738-626-4186	Sequence 4186, Ap
41	7	2.5	324	15	US-10-424-599-204076	Sequence 204076, A
42	7	2.5	325	9	US-09-931-457A-32	Sequence 32, Appl
43	7	2.5	325	15	US-10-424-599-273302	Sequence 273302, A
44	7	2.5	331	15	US-10-369-493-4657	Sequence 4657, Ap
45	7	2.5	331	15	US-10-369-493-7415	Sequence 7415, Ap

ALIGNMENTS

RESULT 1
US-10-108-260A-3952
; Sequence 3952, Application US/10108260A
; Publication No. US20040005560A1
; GENERAL INFORMATION:
; APPLICANT: HELIX RESEARCH INSTITUTE
; TITLE OF INVENTION: NO. US20040005560A1el full length cDNA
; FILE REFERENCE: H1-A0106
; CURRENT APPLICATION NUMBER: US/10/108.260A
; CURRENT FILING DATE: 2002-03-27
; NUMBER OF SEQ ID NOS: 5458
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3952
; LENGTH: 108
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-108-260A-3952

Query Match 16.7%; Score 46; DB 15; Length 108;
Best Local Similarity 100.0%; Pred. No. 8.5e-36;
Matches 46; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 19 KENCRPDGRELGEFRRTTWNIGSISTADGSAVLKLGNTTVICGVKA 64
Db 56 KENCRPDGRELGEFRRTTWNIGSISTADGSAVLKLGNTTVICGVKA 101

RESULT 2
US-09-815-242-5460
; Sequence 5460, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel

; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5460
; LENGTH: 302
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-09-815-242-5460

Query Match 2.9%; Score 8; DB 9; Length 302;
Best Local Similarity 100.0%; Pred. No. 51;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 155 ALLAAAKN 162
DB 72 ALLAAAKN 79

RESULT 3
US-10-138-701-16
; Sequence 16, Application US/10138701
; Publication No. US20030186364A1
; GENERAL INFORMATION:
; APPLICANT: Human Genome Sciences, Inc. et al.
; FILE REFERENCE: PB484
; CURRENT APPLICATION NUMBER: US/10/138,701
; CURRENT FILING DATE: 2002-05-06
; PRIOR APPLICATION NUMBER: US/09/512,255A
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/098,964
; PRIOR FILING DATE: 1998-09-01
; PRIOR APPLICATION NUMBER: US 60/009,861
; PRIOR FILING DATE: 1996-01-05
; PRIOR APPLICATION NUMBER: PCT/ US99/19726
; PRIOR FILING DATE: 1999-08-31
; PRIOR APPLICATION NUMBER: US 08/956,171
; PRIOR FILING DATE: 1997-10-20
; NUMBER OF SEQ ID NOS: 61
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 16
; LENGTH: 308
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-10-138-701-16

Query Match 2.9%; Score 8; DB 14; Length 308;
Best Local Similarity 100.0%; Pred. No. 52;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 155 ALLAAAKN 162

DB 72 ALLAAAKN 79

RESULT 4
US-10-282-122A-43948
; Sequence 43948, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Wall, Daniel
; APPLICANT: Zyskind, Judith
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 43948
; LENGTH: 308
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-10-282-122A-43948

Query Match 2.9%; Score 8; DB 15; Length 308;
Best Local Similarity 100.0%; Pred. No. 52;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 155 ALLAAAKN 162
DB 72 ALLAAAKN 79

RESULT 5
US-10-857-625-624
; Sequence 624, Application US/10857625
; Publication No. US20050026189A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; TITLE OF INVENTION: MICROBIAL OPERONS
; FILE REFERENCE: ELITRA.036A
; CURRENT APPLICATION NUMBER: US/10/857,625

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 23, 2005, 12:33:04 ; Search time 48 Seconds
(without alignments)
429.232 Million cell updates/sec

Title: US-10-736-892-12
Perfect score: 276
Sequence: 1 MAAGFKTVPELYRRFLKE.....TRHKEVSKLDEVIQSMKHK 276

Scoring table: Gapop 60.0 , Gapext 60.0

Searched: 513545 seqs, 74649064 residues

Word size : 0

Total number of hits satisfying chosen parameters: 513545

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Issued Patents AA:*
1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/1/iaa/PTUS_COMB.pep:*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	8	2.9	308	4	US-10-138-701-16
2	7	2.5	87	4	US-09-270-767-38866
3	7	2.5	87	4	US-09-270-767-54083
4	7	2.5	140	4	US-10-141-645-66
5	7	2.5	189	4	US-09-252-991A-17162
6	7	2.5	191	4	US-09-252-991A-24515
7	7	2.5	308	2	US-08-789-609A-2
8	7	2.5	308	3	US-09-108-517-2
9	7	2.5	308	4	US-09-489-039A-7594
10	7	2.5	332	4	US-09-454-279-20
11	7	2.5	335	3	US-09-198-955A-10
12	7	2.5	335	3	US-09-184-217-1
13	7	2.5	335	3	US-09-694-531-10
14	7	2.5	335	4	US-09-789-266-1
15	7	2.5	335	4	US-10-072-152-10
16	7	2.5	389	4	US-09-248-796A-18385
17	7	2.5	404	4	US-09-328-352-6854
18	7	2.5	416	4	US-09-489-039A-13278
19	7	2.5	450	4	US-09-252-991A-25019
20	7	2.5	501	4	US-09-902-540-12135
21	7	2.5	589	4	US-09-252-991A-18093
22	7	2.5	596	4	US-09-902-540-10121
23	7	2.5	663	4	US-09-252-991A-22138
24	7	2.5	666	4	US-09-252-991A-17462
25	7	2.5	689	4	US-09-949-016-11276
26	7	2.5	764	4	US-09-252-991A-32697
27	7	2.5	850	4	US-09-949-016-11324

28	7	2.5	1027	4	US-09-252-991A-17886	Sequence 17886, A
29	7	2.5	1037	4	US-09-340-620A-55	Sequence 55, Appl
30	7	2.5	1049	4	US-09-252-991A-19919	Sequence 19919, A
31	7	2.5	1114	4	US-09-252-991A-24965	Sequence 24965, A
32	7	2.5	1912	4	US-09-495-714C-2	Sequence 2, Appli
33	7	2.5	1977	4	US-09-495-714C-4	Sequence 4, Appli
34	7	2.5	1985	4	US-09-495-714C-6	Sequence 6, Appli
35	6	2.2	9	3	US-09-041-886-45	Sequence 45, Appl
36	6	2.2	11	2	US-08-874-347-11	Sequence 11, Appl
37	6	2.2	11	3	US-09-093-522-11	Sequence 11, Appl
38	6	2.2	11	3	US-09-041-886-43	Sequence 43, Appl
39	6	2.2	12	2	US-08-874-347-17	Sequence 17, Appl
40	6	2.2	12	3	US-09-093-522-17	Sequence 17, Appl
41	6	2.2	13	4	US-08-839-131A-1	Sequence 1, Appli
42	6	2.2	13	4	US-08-839-131A-2	Sequence 2, Appli
43	6	2.2	13	4	US-08-839-131A-3	Sequence 3, Appli
44	6	2.2	14	3	US-09-041-886-3	Sequence 3, Appli
45	6	2.2	14	3	US-09-041-886-4	Sequence 4, Appli

ALIGNMENTS

RESULT 1
US-10-138-701-16
; Sequence 16, Application US/10138701
; Patent No. 6753149
; GENERAL INFORMATION:
; APPLICANT: Human Genome Sciences, Inc. et al.
; TITLE OF INVENTION: Staphylococcus aureus genes and polypeptides
; FILE REFERENCE: PB484
; CURRENT APPLICATION NUMBER: US/10/138,701
; CURRENT FILING DATE: 2002-05-06
; PRIOR APPLICATION NUMBER: US/09/512,255A
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/098,964
; PRIOR FILING DATE: 1998-09-01
; PRIOR APPLICATION NUMBER: US 60/009,861
; PRIOR FILING DATE: 1996-01-05
; PRIOR APPLICATION NUMBER: PCT/ US99/19726
; PRIOR FILING DATE: 1999-08-31
; PRIOR APPLICATION NUMBER: US 08/956,171
; PRIOR FILING DATE: 1997-10-20
; NUMBER OF SEQ ID NOS: 61
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 16
; LENGTH: 308
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-10-138-701-16

Query Match 2.9%; Score 8; DB 4; Length 308;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 155 ALLAAAKN 162
| | | | |
Db 72 ALLAAAKN 79

RESULT 2
US-09-270-767-38866
; Sequence 38866, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 38866

; LENGTH: 87
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
; FEATURE:
; OTHER INFORMATION: Xaa means any amino acid
US-09-270-767-38866

Query Match 2.5%; Score 7; DB 4; Length 87;
Best Local Similarity 100.0%; Pred. No. 39;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 87 PLCSSRF 93
Db 23 PLCSSRF 29

RESULT 3
US-09-270-767-54083
; Sequence 54083, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 54083
; LENGTH: 87
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
; FEATURE:
; OTHER INFORMATION: Xaa means any amino acid
US-09-270-767-54083

Query Match 2.5%; Score 7; DB 4; Length 87;
Best Local Similarity 100.0%; Pred. No. 39;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 87 PLCSSRF 93
Db 23 PLCSSRF 29

RESULT 4
US-10-141-645-66
; Sequence 66, Application US/10141645
; Patent No. 6713078
; GENERAL INFORMATION:
; APPLICANT: Robert Lehrer
; APPLICANT: Alan Waring
; APPLICANT: Alexander Cole
; APPLICANT: Teresa Hong
; TITLE OF INVENTION: Retrocyclins - Antiviral and
; TITLE OF INVENTION: Antimicrobial Peptides
; FILE REFERENCE: UCLA-001CIP
; CURRENT APPLICATION NUMBER: US/10/141,645
; CURRENT FILING DATE: 2002-05-06
; PRIOR APPLICATION NUMBER: 60/284,855
; PRIOR FILING DATE: 2001-04-18
; PRIOR APPLICATION NUMBER: Unassigned
; PRIOR FILING DATE: 2002-04-18
; NUMBER OF SEQ ID NOS: 125
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 66
; LENGTH: 140
; TYPE: PRT
; ORGANISM: Orangutan
US-10-141-645-66

Query Match 2.5%; Score 7; DB 4; Length 140;
Best Local Similarity 100.0%; Pred. No. 59;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 153 TFPALAA 159
Db 7 TFPALAA 13

RESULT 5
US-09-252-991A-17162
; Sequence 17162, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 17162
; LENGTH: 189
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-17162

Query Match 2.5%; Score 7; DB 4; Length 189;
Best Local Similarity 100.0%; Pred. No. 76;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 228 RGRQAVL 234
Db 51 RGRQAVL 57

RESULT 6
US-09-252-991A-24515
; Sequence 24515, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 24515
; LENGTH: 191
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-24515

Query Match 2.5%; Score 7; DB 4; Length 191;
Best Local Similarity 100.0%; Pred. No. 77;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 228 RGRQAVL 234
Db 172 RGRQAVL 178

RESULT 7
US-08-789-609A-2
; Sequence 2, Application US/08789609A
; Patent No. 5827689

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 23, 2005, 04:22:33 ; Search time 3111 Seconds
(without alignments)
10130.895 Million cell updates/sec

Title: US-10-736-892-13
Perfect score: 828
Sequence: 1 atggcgccgggttcacaaac.....ttcagagcatgaacacaaa 828

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 34239544 seqs, 19032134700 residues

Total number of hits satisfying chosen parameters: 68479089

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: gb_est1:*
2: gb_est2:*
3: gb_hic:*
4: gb_est3:*
5: gb_est4:*
6: gb_est5:*
7: gb_est6:*
8: gb_gss1:*
9: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	739.4	89.3	943	3 AK009584	Mus muscu
2	714.6	86.3	1012	6 BY709262	BY709262
3	687	83.0	924	4 B1853518	B1853518
4	683	82.5	996	6 BY762333	BY762333
5	671.8	81.1	935	6 CB204353	CB204353
6	663.8	80.2	811	9 AY400223	AY400223
7	659.4	79.6	944	1 AL531894	AL531894
8	659.2	79.6	1025	5 BM904094	BM904094
9	650	78.5	941	3 CR595030	CR595030
10	643.6	77.7	866	7 CN644054	CN644054
11	638.4	77.1	896	3 CR612849	CR612849
12	638	77.1	903	3 CR602904	CR602904
13	634.6	76.6	841	5 BX425474	BX425474
14	629.8	76.1	1007	1 AL558353	AL558353
15	629	76.0	877	1 AL558475	AL558475
16	620.8	75.0	820	2 BE738362	BE738362
17	617.8	74.6	854	4 B1603115	B1603115
18	613.8	74.1	1098	4 BM467670	BM467670
19	611	73.8	880	5 BQ223602	BQ223602
20	611	73.8	888	1 AL558476	AL558476
21	608.8	73.5	872	1 AL531893	AL531893
22	607.2	73.3	819	4 B1152376	B1152376
23	599	72.3	879	7 CN153352	CN153352
24	598.2	72.2	811	9 AY400221	AY400221

25	596	72.0	810	7 CF579802	CF579802
26	594.4	71.8	747	7 CN348606	CN348606
27	592	71.5	651	7 CF977024	CF977024
28	589	71.1	838	7 CK790115	CK790115
29	587.2	70.9	751	6 CB950339	CB950339
30	583.8	70.5	874	1 AJ819699	AJ819699
31	581.6	70.2	774	6 CB326135	CB326135
32	580.8	70.1	888	5 BQ277403	BQ277403
33	580.8	70.1	951	5 BQ224502	BQ224502
34	579.6	70.0	764	7 CO040283	CO040283
35	579	69.9	631	7 CF114473	CF114473
36	576.6	69.6	739	7 CN155464	CN155464
37	572.6	69.2	757	7 CO040516	CO040516
38	570.6	68.9	721	7 CO734601	CO734601
39	570.2	68.9	874	1 AJ817172	AJ817172
40	569.8	68.8	724	4 BM384065	BM384065
41	568.2	68.6	840	6 CD580163	CD580163
42	565	68.2	647	2 AW990173	AW990173
43	563.8	68.1	740	4 BG568455	BG568455
44	562	67.9	1125	2 BE788716	BE788716
45	559.8	67.6	747	4 BG527301	BG527301

ALIGNMENTS

RESULT 1
AC009584
LOCUS
DEFINITION Mus musculus adult male tongue cDNA, RIKEN full-length enriched library, clone:2310032N20 product:unknown EST, full insert sequence.
AC009584
VERSION AK009584.1 GI:12844467
KEYWORDS HTC; CAP trapper.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
1 Carninci, P. and Hayashizaki, Y.
AUTHORS
TITLE High-efficiency full-length cDNA cloning
JOURNAL Meth. Enzymol. 303, 19-44 (1999)
MEDLINE 99279253
PUBMED 10349636
REFERENCE
2 Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
AUTHORS
TITLE Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes
JOURNAL Genome Res. 10 (10), 1617-1630 (2000)
MEDLINE 20499374
PUBMED 11042159
REFERENCE
3 Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P., Konno, H., Akiyama, J., Nishi, K., Kitsuai, T., Tashiro, H., Itoh, M., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watahiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsuura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.
AUTHORS
TITLE RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer
JOURNAL Genome Res. 10 (11), 1757-1771 (2000)
MEDLINE 20530913
PUBMED 11076861
REFERENCE
4 The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium.
AUTHORS
TITLE Functional annotation of a full-length mouse cDNA collection
JOURNAL Nature 409, 685-690 (2001)
MEDLINE 1152376
PUBMED 1152376
REFERENCE
5 The FANTOM Consortium and the RIKEN Genome Exploration Research

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 23, 2005, 05:52:04 ; Search time 546 Seconds
(without alignments)
9206.649 Million cell updates/sec

Title: US-10-736-892-13
Perfect score: 828
Sequence: 1 atggggccgggttcaaac.....ttcagagcatgaacacacaa 828

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 5633728 seqs, 3035525691 residues

Total number of hits satisfying chosen parameters: 11267456

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications NA.*

```

1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq.*
2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq.*
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq.*
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq.*
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq.*
6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq.*
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq.*
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq.*
9: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq.*
10: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq.*
11: /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq.*
12: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq.*
13: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq.*
14: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq.*
15: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq.*
16: /cgn2_6/ptodata/1/pubpna/US10D_PUBCOMB.seq.*
17: /cgn2_6/ptodata/1/pubpna/US10E_PUBCOMB.seq.*
18: /cgn2_6/ptodata/1/pubpna/US10F_PUBCOMB.seq.*
19: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq.*
20: /cgn2_6/ptodata/1/pubpna/US11_NEW_PUB.seq.*
21: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq.*
22: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq.*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	585	70.7	1372	14	US-10-198-846-11485 Sequence 11485, A
2	499.2	60.3	2835	17	US-10-108-260A-1509 Sequence 1509, Ap
3	409.2	49.4	520	17	US-10-242-535A-20164 Sequence 20164, A
4	409.2	49.4	520	17	US-10-085-783A-20164 Sequence 20164, A
5	341.8	41.3	532	17	US-10-242-535A-49800 Sequence 49800, A
6	341.8	41.3	532	17	US-10-085-783A-49800 Sequence 49800, A
C 7	330.4	39.9	1795	18	US-10-357-930-25092 Sequence 25092, A
C 8	276.2	33.4	362	11	US-09-969-034-2669 Sequence 2669, Ap
9	275	33.2	352	17	US-10-242-535A-50191 Sequence 50191, A
10	275	33.2	352	17	US-10-085-783A-50191 Sequence 50191, A
11	253	30.6	451	11	US-09-969-034-2798 Sequence 2798, Ap

12	245.4	29.6	435	17	US-10-242-535A-51089 Sequence 51089, A
13	245.4	29.6	435	17	US-10-085-783A-51089 Sequence 51089, A
C 14	244.4	29.5	384	9	US-09-917-800A-302 Sequence 302, App
C 15	240	29.0	438	10	US-09-873-367C-935 Sequence 935, App
C 16	240	29.0	438	19	US-10-843-641A-935 Sequence 935, App
17	188.4	22.6	501	9	US-09-783-590-7844 Sequence 7844, Ap
18	178.8	21.8	437	14	US-10-060-036-2224 Sequence 2224, Ap
19	118.2	14.3	535	11	US-09-969-034-4175 Sequence 4175, Ap
20	118.2	14.3	869	14	US-10-198-846-2845 Sequence 2845, Ap
C 21	116.2	14.0	651	11	US-09-969-034-3699 Sequence 3699, Ap
C 22	107.6	13.0	70123	17	US-10-087-192-610 Sequence 610, App
C 23	103.8	12.5	1505	13	US-10-424-535A-26388 Sequence 134388, A
24	101.6	12.3	156	17	US-10-242-535A-26368 Sequence 26368, A
C 25	101.6	12.3	156	17	US-10-085-783A-26368 Sequence 26368, A
26	97.8	11.8	1170	17	US-10-425-114-2929 Sequence 2929, Ap
27	95	11.5	1308	17	US-10-425-114-3139 Sequence 3139, Ap
28	95	11.5	1371	18	US-10-425-115-43059 Sequence 43059, A
29	95	11.5	1379	17	US-10-425-114-32338 Sequence 32338, A
C 30	79.2	9.6	93	14	US-10-060-036-2061 Sequence 2061, Ap
31	78.6	9.5	1106	18	US-10-437-963-17351 Sequence 17351, A
C 32	78.6	9.5	1284	17	US-10-424-535A-134389 Sequence 134389, A
33	68.8	8.3	1148	16	US-10-119-428-54 Sequence 54, Appl
34	68.2	8.2	1776	17	US-10-425-114-25528 Sequence 25528, A
35	68.2	8.2	1842	18	US-10-425-115-178931 Sequence 178931, A
36	66.6	8.0	656	18	US-10-425-115-109522 Sequence 109522, A
37	65	7.9	560	11	US-09-969-034-2950 Sequence 2950, Ap
38	60	7.2	567	13	US-10-027-632-91236 Sequence 91236, A
39	60	7.2	567	13	US-10-027-632-317647 Sequence 317647, A
40	60	7.2	567	17	US-10-027-632-91236 Sequence 91236, A
41	60	7.2	567	17	US-10-027-632-317647 Sequence 317647, A
42	59.6	7.2	1208	18	US-10-739-930-417 Sequence 417, App
43	58.6	7.1	675	18	US-10-767-701-7418 Sequence 7418, Ap
44	56.2	6.8	498	18	US-10-767-701-25714 Sequence 25714, A
45	56	6.8	448	18	US-10-767-701-30731 Sequence 30731, A

ALIGNMENTS

RESULT 1

US-10-198-846-11485/c
; Sequence 11485, Application US/10198846
; Publication No. US2003009974A1
; GENERAL INFORMATION:
; APPLICANT: Lillie, James
; APPLICANT: Xu, Yongyao
; APPLICANT: Wang, Youzhen
; APPLICANT: Steimann, Kathleen
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS
; TITLE OF INVENTION: FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; TITLE OF INVENTION: THERAPY OF BREAST CANCER
; FILE REFERENCE: MRI-049
; CURRENT APPLICATION NUMBER: US/10/198,846
; CURRENT FILING DATE: 2002-07-18
; PRIOR APPLICATION NUMBER: 60/306,220
; PRIOR FILING DATE: 2001-07-18
; NUMBER OF SEQ ID NOS: 14084
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11485
; LENGTH: 1372
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-198-846-11485

Query Match 70.7%; Score 585; DB 14; Length 1372;
Best Local Similarity 89.4%; Pred. No. 5.6e-167;
Matches 641; Conservative 0; Mismatches 75; Indels 1; Gaps 1;
QY 1 ATGGGGCCGGGTTCACAACTGTGGAAACCGCTGAGTATTACAGGAGATTCTTGAAGAA 60
1278 ATGGGCGCTGGGTTCACAAACCGTGGAACTCTGGAGTATTACAGGAGATTCTTGAAGAG 1219
QY 61 AACTGCCGTCAGATGGAAGAGAACTTGGTGAATTTCAGAACCACTGTCAACATAGGT 120

Db 1218 AACTGCCGCTCTGATGGAAGAGAACTTGGTGAATTCAGAACCAACTGTCAACATCGGT 1159
Qy 121 TCGATCAGTACAGCGGATGGCTCTGCTCTAGTGAAGCTGGGACACCAACAGTCATTTGT 180
Db 1158 TCAATTAAGTACCGCAGATGGTCTGCTTGTAGTGAAGTTGGGAAATATCTACAGTAATCTGT 1099
Qy 181 GGAGTTAAAGCAGAAATTTGAGCAGCACCACTAGATGCTCCCTGTAGAGGATATGTCGTC 240
Db 1098 GGAGTTAAAGCAGAAATTTGAGCAGCACCACTAGATGCTCCCTGTAGAGGATATGTCGTC 1039
Qy 241 CCTAATGTGGACCTACCAACCGCTGTGTCATCGAGGTTTCGAGTGAACCTCTCTGGAGAA 300
Db 1038 CCTAATGTGGATCTACCAACCGCTGTGTCATCGAGATTCGCGTCTGGACCTCTCTGGAGAA 979
Qy 301 GAGGCTCAAGTAAACGCGCAGTTCAATTCGAGATGTCATTCGAGACTCAGACATATTAAG 360
Db 978 GAGGCTCAAGTAAACGCGCAGTTCAATTCGAGATGTCATTCGAGACTCAGACATATTAAG 919
Qy 361 AAAGAGGACTTATGCAATTTCTCCAGGGAAGCTTCTGGGTTCTATACCTGTGACCTTATT 420
Db 918 AAAGAGGACTTATGCAATTTCTCCAGGGAAGCTTCTGGGTTCTATACCTGTGATCTCAT 859
Qy 421 TGCCTAGACTACGATGGGAAACATTTTGGATGCTCTGACATTTGCTTTGTAGAGCTTTA 480
Db 858 TGCCTAGACTACGATGGGAAACATTTTGGATGCTCTGACATTTGCTTTGTAGAGCTTTA 799
Qy 481 AAGATCTACAGTTGCTGCTGAGTTACTATAAATGAAGAACTGCTTTAGCGGAGTCAAT 540
Db 798 AAAAATGTACAGTTGCTGCTGAGTTACTATAAATGAAGAACTGCTTTAGCAGAGTTAAT 739
Qy 541 TTAAGAGAAAGAAAGTATTTGAATGTTAGAGCAACCCAGTTGCTTCTCATTTGCTGTG 600
Db 738 TTAAGAGAAAGAAAGTATTTGAATGTTAGAGCAACCCAGTTGCTTCTCATTTGCTGTG 679
Qy 601 TTTGATGACACTTTGCTGATAGTGCATCTTACCGGGAGAGGGGACCCCTGTCTCAAG 660
Db 678 TTTGATGACACTTTGCTGATAGTGCATCTTACCGGGAGAGGGGACCCCTGTCTCAAG 620
Qy 661 AACCTTAACCTAGTAAATGAGCAGGAGGAGGCAAGCTGCTGCTTCCACAGCCAGG 717
Db 619 AACCTTAACCTAGTAAATGAGCAGGAGGAGGCAAGCTGCTGCTTCCACAGCCAGG 563

RESULT 2

US-10-108-260A-1509
; Sequence 1509, Application US/10108260A
; Publication No. US20040005560A1
; GENERAL INFORMATION:
; APPLICANT: HELIX RESEARCH INSTITUTE
; TITLE OF INVENTION: No. US20040005560A1e1 full length cDNA
; FILE REFERENCE: H1-A0106
; CURRENT APPLICATION NUMBER: US/10/108,260A
; CURRENT FILING DATE: 2002-03-27
; NUMBER OF SEQ ID NOS: 5458
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 1509
; LENGTH: 2835
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-108-260A-1509

Query Match 60.3%; Score 499.2; DB 17; Length 2835;
Best Local Similarity 87.0%; Pred. No. 1.2e-140;
Matches 571; Conservative 0; Mismatches 83; Indels 2; Gaps 2;

Qy 174 CATTTGTGGAGTTAAAGCAGAAATTTGAGCAGCACCACTAGATAGATCCCTCTGATAGAGATA 233
Db 2079 CATTTCTTTGCTACTATAGGAATTTGAGCAGCACCACTAGATAGATCCCTCTGATAGAGATA 2138
Qy 234 TGTGCTCCCTAAATGTGAGCTTACCAACCGCTGTGTTTCATCGAGGTTTCGAGTGCACCTCC 293
Db 2139 CGTTGTTCTTAATGTGAGCTTACCAACCGCTGTGTTTCATCGAGGTTTCGAGTGCACCTCC 2198

Qy 294 TCGAGAAGAGGCTCAAGTAAACAGCCAGTTTCATTGAGATGTTCATTGAGAACTCACAAT 353
Db 2199 TCGAGAAGAGGCTCAAGTAAACAGCCAGTTTCATTGAGATGTTCATTGAGAACTCACAAT 2258
Qy 354 AATTAAAGAAAGAGGACTTATGCAATTTCTCCAGGAAAGCTTGTGGTTCTATACCTGTA 413
Db 2259 AATTCAAGAAAGAGGACTTATGCAATTTCTCCAGGAAAGCTTGTGGTTCTATACCTGTA 2318
Qy 414 CCTATTATGCTAGACTACGATGGGAAACATTTTGGATGCTGCAATTTGCTTTGTAGC 473
Db 2319 TCTCATTTGCTCGACTACGATGGGAAACATTTTGGATGCTGCAATTTGCTTTGTAGC 2378
Qy 474 AGCTTTAAAGAAATGTACAGTTGCTGAAAGTTACTATAAATGAAGAACTGCTTTAGCGGA 533
Db 2379 GGCTTTAAAGAAATGTACAGTTGCTGAAAGTTACTATAAATGAAGAACTGCTTTAGCGGA 2438
Qy 534 AGTCAATTTAAAGAAAGAAAGTTATTGTAATGTTAGAGCAACCCAGTTGCTACTTCAAT 593
Db 2439 AGTTAATTTAAAGAAAGAAAGTTATTGTAATGTTAGAGCAACCCAGTTGCTACTTCAAT 2498
Qy 594 TGCCTGTTTGTAGTACACTTTGCTGATGCTGATGCTGATGCTGATGCTGATGCTGATGCTG 653
Db 2499 TGCCTGTTTGTAGTACACTTTGCTGATGCTGATGCTGATGCTGATGCTGATGCTGATGCT 2557
Qy 654 CCACAGGAACCTTTAAACCGTAGTAATGAGCAGGAGGAGCAAGCTGCTGCTTCCACAAGC 713
Db 2558 CAACAGGAACCTTTAAACCGTAGTAATGAGCAGGAGGAGCAAGCTGCTGCTTCCACAAGC 2617
Qy 714 CAGGTGGAGTGGGCT-GCTGGAGCTAAACTTCAAGTGCATGAGTGCAGTGCAGTGCAGTGCAG 772
Db 2618 CAGGTGGAGTGGGCTAACTTCAAGTGCATGAGTGCAGTGCAGTGCAGTGCAGTGCAGTGCAG 2677
Qy 773 GACACAAAGAGTGAGCAAACTACTGATGAGTAAATTCAGAGCATGAAACACAAA 828
Db 2678 GACACAAAGAGTGAGCAAACTACTGATGAGTAAATTCAGAGCATGAAACACAAA 2733

RESULT 3

US-10-242-535A-20164
; Sequence 20164, Application US/10242535A
; Publication No. US20040013663A1
; GENERAL INFORMATION:
; APPLICANT: ChondroGene Inc.
; APPLICANT: Liaw, C.C.

; TITLE OF INVENTION: Compositions and Methods Relating to Osteoarthritis
; FILE REFERENCE: 4231/2005
; CURRENT APPLICATION NUMBER: US/10/242,535A
; CURRENT FILING DATE: 2002-09-12
; PRIOR APPLICATION NUMBER: US 10/085,783
; PRIOR FILING DATE: 2002-02-28
; PRIOR APPLICATION NUMBER: US 60/305,340
; PRIOR FILING DATE: 2001-07-13
; PRIOR APPLICATION NUMBER: US 60/275,017
; PRIOR FILING DATE: 2001-03-12
; PRIOR APPLICATION NUMBER: US 60/271,955
; PRIOR FILING DATE: 2001-02-28
; NUMBER OF SEQ ID NOS: 58994
; SOFTWARE: Patent in version 3.2
; SEQ ID NO 20164
; LENGTH: 520
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (24)..(24)
; OTHER INFORMATION: n is a, c, g, or t
US-10-242-535A-20164

Query Match 49.4%; Score 409.2; DB 17; Length 520;
Best Local Similarity 90.0%; Pred. No. 1.1e-113;
Matches 460; Conservative 0; Mismatches 49; Indels 2; Gaps 2;

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 23, 2005, 04:33:09 ; Search time 189 Seconds
(without alignments)
7168.450 Million cell updates/sec

Title: US-10-736-892-13
Perfect score: 828
Sequence: 1 atggcgccgggttcataaac.....ttcagagcatgaacacacaaa 828

Scoring table: IDENTITY NUC
Gapop 10_0 , Gapext 1.0

Searched: 1202784 seqs, 818138359 residues

Total number of hits satisfying chosen parameters: 2405568

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents NA:
1: /cgn2_6/prodata/1/ina/5A COMB.seq.*
2: /cgn2_6/prodata/1/ina/5B COMB.seq.*
3: /cgn2_6/prodata/1/ina/6A COMB.seq.*
4: /cgn2_6/prodata/1/ina/6B COMB.seq.*
5: /cgn2_6/prodata/1/ina/PCTUS COMB.seq.*
6: /cgn2_6/prodata/1/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	621.2	75.0	3887	4	US-09-976-594-660
C 2	68.8	8.3	1058	4	US-09-976-594-325
C 3	43.2	5.2	7218	1	US-08-232-463-14
C 4	37.2	4.5	813	1	US-07-901-707-11
C 5	37.2	4.5	813	1	US-07-901-707-57
C 6	37.2	4.5	813	1	US-07-988-430-11
C 7	37.2	4.5	813	1	US-07-988-430-57
C 8	37.2	4.5	813	1	US-08-425-336-11
C 9	37.2	4.5	813	1	US-08-488-113B-11
C 10	37.2	4.5	813	1	US-07-477-484B-11
C 11	37.2	4.5	813	2	US-08-646-360-11
C 12	37.2	4.5	813	2	US-08-621-803-246
C 13	37.2	4.5	813	3	US-08-839-765-11
C 14	37.2	4.5	813	3	US-09-136-389-11
C 15	37.2	4.5	813	3	US-09-217-352-246
C 16	37.2	4.5	813	3	US-09-610-838-11
C 17	37.2	4.5	813	3	US-09-711-485-11
C 18	37.2	4.5	813	5	PCT-US92-09487-11
C 19	37.2	4.5	813	5	PCT-US92-09487-57
C 20	37.2	4.5	955	2	US-08-621-803-258
C 21	37.2	4.5	955	3	US-08-217-352-258
C 22	37.2	4.5	1003	2	US-08-621-803-252
C 23	37.2	4.5	1003	3	US-09-217-352-252
C 24	37.2	4.5	1072	2	US-08-621-803-250
C 25	37.2	4.5	1072	3	US-09-217-352-250
C 26	37.2	4.5	92304	4	US-09-949-016-15943
C 27	37	4.5	601	4	US-09-949-016-155573

C 28	37	4.5	601	4	US-09-949-016-155672	Sequence 155672, A
C 29	37	4.5	93971	4	US-09-949-016-16097	Sequence 16097, A
C 30	37	4.5	93971	4	US-09-949-016-16098	Sequence 16098, A
C 31	36.4	4.4	1830121	4	US-09-557-884-1	Sequence 1, Appli
C 32	36.4	4.4	1830121	4	US-09-643-990A-1	Sequence 1, Appli
C 33	36.2	4.4	5137	4	US-08-956-171E-217	Sequence 217, App
C 34	36.2	4.4	5137	4	US-08-781-986A-217	Sequence 217, App
C 35	35.8	4.3	601	4	US-09-949-016-96261	Sequence 96261, A
C 36	35.8	4.3	601	4	US-09-949-016-96262	Sequence 96262, A
C 37	35.8	4.3	387902	4	US-09-949-016-14543	Sequence 14543, A
C 38	35.8	4.3	421883	4	US-09-949-016-12557	Sequence 12557, A
C 39	35.4	4.3	601	4	US-09-949-016-96260	Sequence 96260, A
C 40	35.4	4.3	832	4	US-09-621-976-2813	Sequence 2813, Ap
C 41	35.2	4.3	1141	4	US-09-806-708B-22	Sequence 22, Appl
C 42	35.2	4.3	239527	4	US-09-949-016-15980	Sequence 15980, A
C 43	35	4.2	832	4	US-09-621-976-2813	Sequence 2813, Ap
C 44	35	4.2	2069	4	US-09-270-767-12077	Sequence 12077, A
C 45	34.8	4.2	951	4	US-09-107-532A-1037	Sequence 1037, Ap

ALIGNMENTS

RESULT 1
US-09-976-594-660/c
; Sequence 660, Application US/09976594
; Patent No. 6673549
; GENERAL INFORMATION:
; APPLICANT: Furness, Michael
; APPLICANT: Buchbinder, Jenny
; TITLE OF INVENTION: GENES EXPRESSED IN C3A LIVER CELL CULTURES TREATED WITH STEROIDS
; FILE REFERENCE: PA-0041 US
; CURRENT APPLICATION NUMBER: US/09/976,594
; CURRENT FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: 60/240,409
; PRIOR FILING DATE: 2000-10-12
; NUMBER OF SEQ ID NOS: 1143
; SOFTWARE: PERL Program
; SEQ ID NO 660
; LENGTH: 3887
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. 6673549 033627.33
; NAME/KEY: unsure
; LOCATION: 2483, 2486
; OTHER INFORMATION: a, t, c, g, or other
US-09-976-594-660

Query Match	75.0%	Score 621.2;	DB 4;	Length 3887;
Best Local Similarity	86.6%;	Pred. No. 2.2e-200;		
Matches 729;	Conservative 0;	Mismatches 88;	Indels 25;	Gaps 3;
QY	11	GGTTCAAACTGTGGAAACCGCTGGAGTATTACAGGAGATTTCTGAAAGAAAACGCGTC	70	
Db	3793	GTTCAAAACCGTGAACCTCTGGAGTATTACAGGAGATTTCTGAAAGAGAACTGCGTC	3734	
QY	71	CAGATGGAGAGAACTTTGGTGAATTCAGAACACCACTGTCAACATAGTTCCATCAGTA	130	
Db	3733	CTGATGGAGAGAACTTTGGTGAATTCAGAACCACTGTCAACATAGTTCCATCAGTA	3674	
QY	131	CAGCGGATGGCTCTGCTCTAGTGAACCTGGGAAACCAACAGTCATTTGTGGAGTTAAAG	190	
Db	3673	CCGAGATGGTCTGCTTTAGTGAATTCGGAATACTACAGTAATCTGTGGAGTTAAAG	3614	
QY	191	CA-----GAATTCAGCACCAACAGTAGATGCCCTCATAG	227	
Db	3613	CAATGACTTTCAGTAACCTGTCCAGGAATTTGCAGCAACCATCAACAGATGCCCTCATAG	3554	
QY	228	AGGATATGTCGTCCTTAATGTGGACCTACACCGCTGTGTTTCATCGAGGTTTCGGAATGG	287	
Db	3553	AGGATACGTTGTTCTCTTAATGTGGATCTACACCCCTGTGTTTCATCGAGATTCGGGTTGG	3494	

QY 288 ACCTCCTGGAGAAGAGGCTCAAGTAACAGCCAGTTCATTGACAGATGTCATTGAGAACTC 347
Db 3493 ACCTCCTGGAGAAGAGGCTCAAGTGGCTAGCCAGTTCATTGACAGATGTCATTGAGAACTC 3434
QY 348 ACACATAAATAAGAAAGAGGCTTATGATGCTTCCAGGGAAGCTTGGTTCCTATATA 407
Db 3433 ACAGATAAATCAGAAAGAGGACTTATGATGCTTCCAGGGAAGCTTGGTTCCTATATA 3374
QY 408 CTGTGACCTTATTTGCTTACCTAGACTAGGAGGAAATTTTGGATGCTGACATTTGCTTT 467
Db 3373 CTGTGATCTCAATTTGCTTACCTAGACTAGGAGGAAATTTTGGATGCTGACATTTGCTTT 3314
QY 468 GTTAGCAGCTTAAAGAAATGATGATGCTTGAAGTACTATAAATGAAGAAAGCTGCTTT 527
Db 3313 CTGTAGCGCTTTAAAGAAATGATGATGCTTGAAGTACTATAAATGAAGAAAGCTGCTTT 3254
QY 528 AGCGGAAGTCAATTTAAAGAAAGAAAGTATTTGAATGTTAGACAAACCCAGTGTCTAC 587
Db 3253 AGCAGAAGTAAATTTAAAGAAAGAAAGTATTTGAATATTTAGAACTCATCCAGTTGCAAC 3194
QY 588 TTCAATTTGCTGCTTGTGATGACACTTTGCTGATGCTGATGCTTACCGGGAGGAGGGCA 647
Db 3193 TTCTTTGCTGCTTGTGATGACACTTTGCTTATAGTGTGACCTTACTGGAGAGGAGGAACA 3134
QY 648 CCCTGTCCACAGGAACCTTAAACCGTAGTAAATGGACGAGGAAGCAAGCTGTGCTGCTTC 707
Db 3133 -TCTGGCAACAGGAACCTTAAACATAGTAAATGGATGAGGAAGCAACTGTGCTGCTTC 3075
QY 708 ACAAGCCAGTGGAGTGGCT-GCTGGAGTAAACTTACAGACTGATGATGATGAGCAG 766
Db 3074 ACAAAACAGTGGAGTGGCTTAACTGGAGCTAACTTACAGACTGATGATGAGCAGCAG 3015
QY 767 TAACGAGACAAAGAACTGAGCAACTTACTGGATGAATTAATTCAGAGCATGAAACACA 826
Db 3014 TTAACGAGACAAAGAACTGAGCAACTTACTGGATGAATTAATTAAGAGTATGAAACCCA 2955
QY 827 AA 828
Db 2954 AA 2953

RESULT 2

US-09-976-594-325
; Sequence 325, Application US/09976594
; Patent No. 6673549
; GENERAL INFORMATION:
; APPLICANT: Furness, Michael
; APPLICANT: Buchbinder, Jenny
; TITLE OF INVENTION: GENES EXPRESSED IN C3A LIVER CELL CULTURES TREATED WITH STEROIDS
; FILE REFERENCE: PA-0041 US
; CURRENT APPLICATION NUMBER: US/09/976,594
; CURRENT FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: 60/240,409
; PRIOR FILING DATE: 2000-10-12
; NUMBER OF SEQ ID NOS: 1143
; SOFTWARE: PERL Program
; SEQ ID NO 325
; LENGTH: 1058
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. 6673549 2551987CBI
US-09-976-594-325

Query Match 8.3%; Score 68.8; DB 4; Length 1058;
Best Local Similarity 46.0%; Pred. No. 1.8e-12;
Matches 232; Conservative 0; Mismatches 272; Indels 0; Gaps 0;
QY 1 ATGGCGCGGGTTCAAAACCTGAGACCGCTGGAGTATACAGAGATTTCTGAAGAA 60
Db 36 ATGGCGTCCGTGACGCTGAGCGGAGGAGGAGGTGTACATCGTGCATGGCGTCCAGAA 95

QY 61 AACTGCGCTCAGATGAAGAGAACTTGGTGAATTCAGAACTGCAACCACTGTCAACATAGGT 120
Db 96 GACCTCCGTGTGATGAGCCGTGGCTGTGAGGACTACCGATGTGTCGAAGTGGAACTGAT 155
QY 121 TCGATCAGTACAGCGGATGCTCTGTCTTAGTGAAGCTGGGGAAACACACAGTCAATTTGT 180
Db 156 GTGTGTCCAACACTAGTGGGTCCGCCAGGCTCAAGCTGGGTACACACAGACATCTTGGTG 215
QY 181 GGAGTTAAAGCAGAAATTTGAGCAGCACACAGTATAGTCCCTGTATAGAGATATGTGTC 240
Db 216 GGAGTGAAGACAGAAATGGGGAGCGCCGAAGCTGAGAGAAACAAATGAAGGCTACTTTGGAG 275
QY 241 CCTAATCTGACCTTACCACCGCTGTGTTCATCGAGTTTCGGACTGAGACCTCTCTGGAGAA 300
Db 276 TTCTTTGTGCTGTTGTCAGCAGTGTACCCCTGAAATTTGAAGTAGAGAGGTGATGAC 335
QY 301 GAGGCTCAAGTAACACCGCAGTTTCAATGCAAGATGTCAATGAGAACTCACATAAATTAAG 360
Db 336 CTGGCACCAGATCGCTAACACCCCTATCGGATATTTAAACAATAAAGCAGTGTGAC 395
QY 361 AAAGAGACTTATGCAATTTCTCCAGGGAAGCTTGGTCTTATATCTGTGACCTTATT 420
Db 396 TTAAGAGCCCTCTGCAATTAGTCTCGGAGCACCTGCTGGGTCTCTATATGTGATGTGCTG 455
QY 421 TGCTAGACTACGATGGAACATTTTGGATGCTGCACTTTGCTTTGTTAGCAGCTTTA 480
Db 456 CTCTGGAATGTGTGGAATTTGTTGATGCCATTTCCATTTGCTGTAAGGCTGCTCTC 515
QY 481 AAGAATGTACAGTTGCTGCAAGTT 504
Db 516 TTCAATACAAGGATACCAAGGTT 539

RESULT 3

US-08-232-463-14/c
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHEIFLINGER, F.
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/232,463
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/935,313
; FILING DATE:
; APPLICATION NUMBER: EP 91 114 300.6
; FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 30472/114 IMMU
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)836-9300
; TELEFAX: (703)683-4109
; TELEX: 899149

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 23, 2005, 03:15:14 ; Search time 521 Seconds
(without alignments)
9407.957 Million cell updates/sec

Title: US-10-736-892-13
Perfect score: 828
Sequence: 1 atggggcggtttcaaac.....ttcagagcatgaacacacaa 828

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 4390206 seqs, 2959870667 residues

Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_16Dec04:*
1: Geneseqn1980s:*
2: Geneseqn1990s:*
3: Geneseqn2000s:*
4: Geneseqn2001as:*
5: Geneseqn2001bs:*
6: Geneseqn2002as:*
7: Geneseqn2002bs:*
8: Geneseqn2003as:*
9: Geneseqn2003bs:*
10: Geneseqn2003cs:*
11: Geneseqn2003ds:*
12: Geneseqn2004as:*
13: Geneseqn2004bs:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	664.2	80.2	1009	6	ABN59828 Novel hum
2	664.2	80.2	2161	13	ADP55134 Human PRO
3	637.4	77.0	933	4	AAF22429 Human bre
C 4	621.2	75.0	3887	12	ADL12931 Human ste
5	586.6	70.8	811	13	ACN39626 Tumour-as
C 6	585	70.7	1372	11	ACN90335 Breast ca
7	499.2	60.3	2835	11	ADM02824 Human cdn
8	442	53.4	636	4	AAF22498 Human bre
9	442	53.4	636	4	AAF22591 Human bre
10	433.4	52.3	2428	9	AAI57565 Human Opa
C 11	399.2	48.2	555	13	ADQ56453 Novel can
C 12	354.2	42.8	687	4	AAF22499 Human bre
C 13	354.2	42.8	687	4	AAF22592 Human bre
C 14	330.4	39.9	1795	5	ABV25103 Human pro
15	320	38.6	2189	13	ADR06542 Full leng
C 16	276.2	33.4	362	6	ABQ58974 Human col
C 17	253	30.6	451	6	ABQ59103 Human col
C 18	244.4	29.5	384	6	ABK62395 Rat seque
C 19	244.4	29.5	384	10	ADB50301 Primary r
C 20	244.4	29.5	384	12	ADP71722 Renal tox

C 21	240	29.0	438	6	ABL62598 Colon ade
22	178.8	21.6	437	6	ABV96816 Human pan
23	118.2	14.3	263	4	AAI11494 Human bre
24	118.2	14.3	534	6	ABQ60480 Human col
25	118.2	14.3	869	11	ACN81695 Breast ca
26	117.2	14.2	1061	3	AAC42258 Arabidops
C 27	116.2	14.0	651	6	ABQ60004 Human col
C 28	107.6	13.0	70123	11	ACN44254 Human gen
29	90	10.9	349980	5	AAH41223 Pyrococcu
C 30	88	10.6	110000	12	Continuation (15 o
31	88	10.6	110000	12	Continuation (7 of
C 32	88	10.6	110000	12	Continuation (15 o
33	88	10.6	110000	12	Continuation (7 of
C 34	88	10.6	110000	12	Continuation (15 o
35	88	10.6	110000	12	Continuation (7 of
C 36	79.2	9.6	93	6	ABV96653 Human pan
37	68.8	8.3	1044	10	ADC30625 Human nov
38	68.8	8.3	1045	13	ADP55025 Human PRO
39	68.8	8.3	1047	13	ACN37914 Tumour-as
40	68.8	8.3	1058	12	ADL12596 Human ste
C 41	65.2	7.9	1170	10	ADC32435 Human nov
42	65.2	7.9	110000	11	ADM27081_03
C 43	65	7.9	560	6	ABQ59255 Human col
44	59.6	7.2	1140	3	AAC50582 Arabidops
45	59.6	7.2	1143	3	AAC39970 Arabidops

ALIGNMENTS

RESULT 1
ABN59828
ID ABN59828 standard; cdna; 1009 BP.
XX
AC ABN59828;
XX
DT 28-JUN-2002 (first entry)
DE
DS Novel human coding sequence SEQ ID NO: 239.

XX Human; antianaemic; vulnery; antiinflammatory; immunomodulator;
XX antiinfertility; cerebroprotective; cytostatic; rheumatic; gene therapy;
XX neuroprotective; antiparkinsonian; protein therapy; EST;
XX expressed sequence tag; gene; ss.
XX Homo sapiens.
XX WO200222660-A2.
XX
XX 21-MAR-2002.
XX
XX 10-SEP-2001; 2001WO-US026015.
XX
XX 11-SEP-2000; 2000US-00659671.
XX (HYSE-) HYSEQ INC.

XX Tang YT, Liu C, Zhou P, Asundi V, Zhang J, Zhao QA, Ren F;
XX Xue AJ, Yang Y, Wehrman T, Drmanac RT;
XX WPI; 2002-292408/33.
XX P-PSDB; ABB97415.
XX
XX An isolated polynucleotide for treating diseases associated with its
XX encoded polypeptide such as cancer and multiple sclerosis.

XX Claim 1; SEQ ID NO 239; 509pp; English.
XX
XX The present invention provides the protein and coding sequences of 444
XX novel human proteins. These were isolated from expressed sequences tags
XX (ESTs). They can be used to stimulate cell growth, to regulate
XX haematopoiesis e.g. to treat aplastic anaemia, to help tissue regrowth
XX e.g. in burn treatment, to regulate the immune system e.g. to treat

CC multiple sclerosis, to regulate activin or inhibit e.g. to treat
CC infertility, to regulate haemostasis or thrombolysis e.g. to treat stroke
CC and cancer, to screen for drugs, to treat inflammatory conditions e.g.
CC rheumatoid arthritis, and to treat nervous system disorders e.g.
CC Parkinson's disease. The present sequence is a coding sequence of the
CC invention

SQ Sequence 1009 BP; 328 A; 199 C; 223 G; 259 T; 0 U; 0 Other;

Query Match 80.2%; Score 664.2; DB 6; Length 1009;
Best Local Similarity 89.1%; Pred. NO. 9e-191;
Matches 739; Conservative 0; Mismatches 88; Indels 2;

Qy 1 ATGCGGGCGGGTTCAAAACTGTGGAACCGCTGGAGTATTACAGGAGATTTCTGAAAGAA 60
|||
Db 77 ATGCGGGTGCGGTTCAAAACCGTGGAACCTCTGGAGTATTACAGGAGATTTCTGAAAGAG 136

Qy 61 AACTGCCGTCAGATGGAGAGA
 |||
Dd 137 AACTGCCGTCCTGATGGAGAGA

Qy	121	TCGATCAGTACAGCGGATGGCTCTGTCTCTAGTAGAAGCTGGGGAAACCACCAAGTCATTGT	180
Db	197	TCAATTAGTACCGCAGATCGTTCTGCTTTAGTGAAAGTTGGGAAATACTACAGTAATCTGT	256

Qy 181 GGAGTTAAAGCAGAAATTTGCAGCAGCACCACAGTAGTGCCTGATAGAGGATATGTCGTC 240

Db 257 GGAGTTAAAGCAGAAATTTGCAGCAGCACCATCAACAGATGCCCTGATAAAGGATACGTTGTT 316

Qy	Db
241 CCTAATGTGGACCTPACCAACGGCTGTGTTTCATCGAGTTTCGGACCTGGACCTCCTGGAGAA	300
317 CCTAATGTGGATCTPACCAACCCCTGTGTTTCATCGAGATTCGGGTCTGGACCTCCTGGAGAA	376

Qy	301	GAGGCTCAAGTAAC	CACGCGCTTCAT	TGCAGATGTCAT	TGAGAAGTCAC	CATATTAAG	360
Db	377	GAGGCCCCAAGTG	GTAGCCAGTTCA	TTCAGATGTCAT	TGAAAAATTCAC <td>GATATTCAG</td> <td>436</td>	GATATTCAG	436

Qy	361	AAAGAGGACTTATGCATTCTTCCAGGGAGCTTGCTGGGTTCTATACTGTGACCTTATT	420
Db	437	AAAGAGGACTTATGCATTCTTCCAGGAAAGCTTGCTGGGTTCTATCTGTGATCTCATTT	496

Qy	421 TGCCTAGACTACGATGGGAACATTTTGGATTCCTGCGACATTGTGCTTGTTAGCAGCTTTA 480
Db	497 TGCCTCGACTACGATGGAACAATTTTGGATTCCTGCGACATTGTGCTTGTTAGCGGCTTTA 556

Qy 481 AAGAAATGTACAGTTGCCCTGAAAGTTACTATATAAATGAAGAAACTGCTTTAGCGGAAGTCAAT 540

Db 557 AAAAAATGTACAGTTGCCCTGAAAGTTACTATATAAATGAAGAAACTGCTTTAGCAGAGAGTTAAT 616

Qy 541 TTTAAAGAGAAAAGTTATTTGAATGTTAGAGCAAAACCCAGTTGCTACTTCATTTTGTGTG 600

Db 617 TTTAAAGAGAAAAGTTATTTGAATGTTAGAGCAAAACCCAGTTGCTACTTCATTTTGTGTG 676

Qy 601 TTTTGATGACACTTTTGCTGATAGTTCGATCTTACCGGGGAGGAGGGACCCCTGTCCACAGG 660

Db 677 TTTTGATGACACTTTTGCTTATAGTTGACCCCTACTGGAGAGGAGGAACA-TCTGGCAACAGG 735

Qy 661 AACCTTAACCGTAGTAATGGACGAGGAAGCAAGTGTGCTGTCTTTCACAGCCAGGTGG 720
|||||
Db 736 AACCTTAACAATAGTAATGGATGAGGAAGGCAAACTCTGTTGTCTTTCACAAAACCAAGGTGG 795
|||||

Qy 721 GAGTGGGCT-GCTGGAGCTAAACTTCAGGACTGCATGAGTCGAGCAGTAAAGGACACAA 779

Db 796 AAGTGGGCTAACTGGAGCTAAACTTCAGGACTGATGAGCGGAGCAGTTACAAGACACAA 855

Qy 780 AGAAGTGACAAACTACTGGATGAAGTAATTCAGAGCATGAAACAAA 828
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 856 AGAAGTTRAAAAAAGCTGATGGATGAAGTAATTAAAGAGTATGAAACC AAA 904
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 2
ADP55134

[illegible]

ADP55134;

18-NOV-2004 (first entry)

Human PRO cDNA sequence SEQ ID NO:11110

human: PRO: immune related disease: inflammatory immune response:

immune response stimulator; anidiarragic; anidysparetic; antidiarrhetic; antidiasthmatic; antidiabetic; antiinflammatory; antipsoriatic; antirheumatic; antithyroid; CNS; dermatological; gastrointestinal; haemostatic; hepatotropic; immunostimulant; immunosuppressive; muscular; nephrotropic; neuroprotective; osteopathic; respiratory; vasotropic; virucide; gene therapy; gene; ss.

Homo sapiens.

WO2004039956-A2.

13-MAY-2004

28-00T-2003: 2003W0-115034381

20-0000-2002. 2003112-01234732D

(CETH) CEMENTECH TNC

Aggarwal S, Clark H, Gurney AL, Schoenfeld J, Williams PM; Wood WI, Wu TD;

WPI; 2004-376182/35.

P-PSDB; ADP55135.

New PRO polynucleotides and polypeptides, useful in useful in diagnosing and treating an immune related disease, e.g. systemic lupus erythematosus, rheumatoid arthritis, diabetes mellitus or asthma and in stimulating an immune response.

Claim 2: SEO ID NO 1110: 3009pp: English.

The present invention describes an isolated PRO nucleic acid (1). Also described: (1) a vector comprising (1); (2) a host cell comprising the vector of (1); (3) a process for producing a PRO polypeptides; (4) an isolated PRO polypeptide; (5) a chimeric molecule comprising the polypeptide of (4) fused to a heterologous amino acid sequence; (6) an antibody which specifically binds to a polypeptide of (4); (7) a composition of matter comprising a polypeptide of (4), an agonist or antagonist of the polypeptide or an antibody that binds to the polypeptide in combination with a carrier; (8) an article of manufacture comprising a container, a label on the container and a composition of matter of (7); (9) a method of treating an immune related disease in a mammal; (10) a method for determining the presence of a PRO polypeptide in a sample suspected of having the polypeptide; (11) a method of diagnosing an immune related disease or an inflammatory immune response in a mammal; (12) a method of identifying a compound that inhibits or mimics the activity of or expression of a gene encoding a PRO polypeptide and (13) a method of stimulating the immune response in a mammal. The PRO sequences have antiallergic, antianaemic, antiarthritic, antiasthmatic, antidiabetic, antiinflammatory, antipsoriatic, antirheumatic, antithyroid, CNS, dermatological, gastrointestinal, haemostatic, hepatotropic, immunostimulant, immunosuppressive, muscular, nephrotropic, neuroprotective, osteopathic, respiratory, vasotropic and virucide activities, and can be used in gene therapy. The nucleic acid (1) and the encoded polypeptides, compositions, kits and methods are useful in diagnosing and treating an immune related disease and in stimulating an immune response. The present sequence represents a human PRO nucleotide sequence from the present invention.

Sequence 2161 BP: 639 A: 498 C: 531 G: 493 T: 0 U: 0 Other:

RESULT 2

ADP55134
Best local similarity 89.1%;
Pred. No. 1.3e-190;
Score 001.2, 00.23;
Exact match 00.28;
Length 2102

ID ADP55234 standard; cDNA; 2161 BP.

Best local similarity 0.719; Rec: NO: 1.00 1.00;
Matches 739; Conservative 0; Mismatches 88; Indels 2; Gaps 2;

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 23, 2005, 04:19:13 ; Search time 3650 Seconds
(without alignments)
10992.033 Million cell updates/sec

Title: US-10-736-892-13
Perfect score: 828
Sequence: 1 atggggcggttcaaac.....ttcagagcatgaacacaaa 828

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 4708233 seqs, 24227607955 residues

Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl.*

- 1: gb_ba.*
- 2: gb_hgt.*
- 3: gb_in.*
- 4: gb_cm.*
- 5: gb_ov.*
- 6: gb_pat.*
- 7: gb_ph.*
- 8: gb_pl.*
- 9: gb_pr.*
- 10: gb_ro.*
- 11: gb_sts.*
- 12: gb_sy.*
- 13: gb_un.*
- 14: gb_vl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	739.4	89.3	912	10	MMU505005
2	737.8	89.1	1316	10	BC059089 Mus muscu
3	724	87.4	926	10	BC067250 Mus muscu
4	699.8	84.5	164637	2	AC142180 Rattus no
5	664.2	80.2	1009	6	AX405824 Sequence
6	662.6	80.0	986	9	BC020773 Homo sapi
7	647.2	78.2	944	9	AF025438 Homo sapi
8	637.4	77.0	933	6	AX053242 Sequence
9	621.2	75.0	3887	6	AR448005 Sequence
10	499.2	60.3	2835	9	AX834385 Sequence
11	499.2	60.3	2835	9	AX096810 Homo sapi
12	489.6	59.1	798	9	AF279901 Homo sapi
13	473.6	57.2	1132	5	BX935619 Gallus ga
14	442	53.4	636	6	AX053311 Sequence
15	442	53.4	636	6	AX053404 Sequence
16	433.4	52.3	2428	9	HSM800526 Homo sapi
17	431.6	52.1	1045	5	BC041271 Xenopus l
18	422	51.0	1007	5	BC080360 Xenopus t
19	420.4	50.8	993	5	CR760459 Xenopus t

C	20	418	50.5	172033	9	AL136131	Human DNA
	21	418	50.5	193332	2	AL355349	Homo sapi
	22	409.2	49.4	520	6	CQ675238	Sequence
	23	383.8	46.4	987	5	AY648725	Danio rer
	24	356.8	43.1	816	6	CQ732109	Sequence
C	25	354.2	42.8	687	6	AX053312	Sequence
C	26	354.2	42.8	687	6	AX053405	Sequence
	27	341.8	41.3	532	6	CQ704874	Sequence
C	28	330.4	39.9	1795	6	CQ493225	Sequence
	29	320	38.6	2189	6	CQ849579	Sequence
	30	320	38.6	2189	9	AK126602	Homo sapi
	31	275	33.2	352	6	CQ705265	Sequence
	32	245.4	29.6	435	6	CQ706163	Sequence
C	33	244.4	29.5	384	6	AX400626	Sequence
C	34	240	29.0	438	6	AX330426	Sequence
	35	221.2	26.7	592	5	BX930480	Gallus ga
C	36	122	14.7	14506	1	AE010257	Pyrococcu
C	37	122	14.7	64283	2	AC105693	Continuation (4 of
C	38	122	14.7	213481	2	AC135446	Rattus no
C	39	122	14.7	253431	2	AC128163	Rattus no
	40	118.2	14.3	263	6	CQ418925	Sequence
	41	117.2	14.2	909	8	BT004614	Arabidops
C	42	114	13.8	216239	10	AC100927	Mus muscu
C	43	107.6	13.0	195032	9	AL138706	Human DNA
	44	106	12.8	163831	2	AC150948	Bos tauru
C	45	106	12.8	319000	1	AP000006	Pyrococcu

ALIGNMENTS

RESULT 1
MMU505005
LOCUS
DEFINITION Mus musculus mRNA for CBP-interacting protein 3 (CIP3 gene).
ACCESSION AJ505005
VERSION AJ505005.1 GI:22208742
KEYWORDS CBP-interacting protein 3; CIP3 gene.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1
AUTHORS Heery,D.M., Harries,J.C., Kindle,K.B., Viskaduraki,M., Matsuda,S., Sheppard,H.M. and Davis,A.
TITLE Diverse nuclear proteins compete with SRC1 for interaction with CREB binding protein
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 912)
AUTHORS Heery,D.M.
TITLE Direct Submission
JOURNAL Submitted (05-AUG-2002) Heery D.M., Biochemistry, University of Leicester, University Road, Leicester LE1 7RH, UNITED KINGDOM

FEATURES
source
1..912
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/tissue_type="embryo"
gene
1..912
/genes="CIP3"
CDS
1..831
/function="putative transcriptional regulator"
/codon_start=1
/product="CBP-interacting protein 3"
/protein_id="CAD43467.1"
/db_xref="GI:22208743"

ADGSAVLKGNITVIGVKAEPAPVPDAPGVVNPVNDLPPLCSRRFTGPPGEA
QVTSQIADVDNSQVKKEDLCISPKLAWLYCDLICDYGNDLLDCTALLAAL
KNQLEPVTINEETALAEVNLKKSYLNRVTNPVTSFAVDDTLTLLVDPTGEEHLS
TGTLVTVDDEGKLCCLHKPGSGLTGAKLQDCMSRAVTHHKVSKLLDEVIQSMRHK

BC059089
BC059089.1 GI:37589849
IMAGE:83038207, completed

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: April 23, 2005, 10:16:15 ; Search time 112 Seconds
(without alignments)
7571.457 Million cell updates/sec

Title: US-10-736-892-13

Perfect score: 1490

Sequence: 1 atggggccgggttcaaac.....ttcagagcatgaacacaaa 828

Scoring table:

BLOSUM62	Xgapop 10.0	Xgapext 0.5
Ygapop 10.0	Ygapext 0.5	
Fgapop 6.0	Fgapext 7.0	
Delop 6.0	Delext 7.0	

Searched: 1612378 seqs, 512079187 residues

Total number of hits satisfying chosen parameters: 3224756

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

```
-MODEL=frame+ n2p.model -DBV=xlp
-O=/cgn2.1/USPTO.spool_p/US10736892/runat.22042005.114830.22435/app.query.fasta_1.967
-DB=Uniprot.03 -Qfmt=fastan -SUPFIX=rup -MINMATCH=0.1 -LOOPEXT=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10736892 @CGN 1.1 152 @runat.22042005.114830.22435 -NCPU=3
-NO MAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DSV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7
```

Database : Uniprot.03.*

1: uniprot.sprot.*

2: uniprot.trembl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1353	90.8	276	1 RR43 MOUSE	Q9d753 mus musculu
2	1348	90.5	276	2 Q6PCW8	Q6pcw8 mus musculu
3	1318	88.5	270	2 Q6NX62	Q6nx62 mus musculu
4	1313	88.1	276	1 RR43 HUMAN	Q96b26 homo sapien
5	1051	70.5	276	2 Q66KK0	Q66kk0 xenopus tro
6	1049	70.4	276	2 Q8AVT6	Q8avt6 xenopus lae
7	1004	67.4	277	2 Q6DRN4	Q6drn4 brachydanio
8	591	39.7	280	2 Q7QDD9	Q7qd9 anopheles g
9	479.5	32.2	270	1 RR43 SCHPO	Q10205 schizosacch
10	463	31.1	302	2 Q9ZUI4	Q9zu14 arabidopsis
11	402.5	27.0	277	1 ECX2 PYRFU	Q8u0m0 pyrococcus
12	384	25.8	274	1 ECX2 PYRHO	O59224 pyrococcus
13	377.5	25.3	274	1 ECX2 PYRAB	Q9v118 pyrococcus
14	367	24.6	271	1 ECX2 METHH	O26778 methanobact
15	354	23.8	257	2 Q64B51	Q64b51 uncultured
16	353	23.7	257	2 Q64V77	Q64v77 uncultured

17	348	23.4	257	2 Q64D54	Q64d54 uncultured
18	342.5	23.0	267	1 ECX2_METKA	Q8tyc2 methanopyru
19	341.5	22.9	276	1 ECX2_AERPE	Q9yc05 aeropyrum p
20	339	22.8	260	2 Q64CA6	Q64ca6 uncultured
21	338.5	22.7	260	2 Q6L224	Q6l224 picrophilus
22	338	22.7	259	1 ECX2_ARCFU	O29756 archaeoglob
23	337	22.6	260	2 Q649Q1	Q649q1 uncultured
24	335	22.5	260	2 Q6MZC0	Q6mzc0 uncultured
25	332.5	22.3	260	1 ECX2_THEAC	Q9hip1 thermoplasm
26	331	22.2	260	1 ECX2_THEVO	Q97b24 thermoplasm
27	331	22.2	275	1 ECX2_SULTO	Q975g9 sulfolobus
28	327	21.9	238	2 Q64DM5	Q64dm5 uncultured
29	325	21.8	275	1 ECX2_SULSO	Q9uxc0 sulfolobus
30	313	21.0	395	2 Q69T30	Q69t30 cryza sativ
31	309	20.7	266	1 ECX2_METMA	Q9ptt7 methanosarc
32	308.5	20.7	274	1 ECX2_PYRAE	Q8zvn0 pyrobaculum
33	307	20.6	266	1 ECX2_METAC	Q8txg5 methanosarc
34	300	20.1	307	2 Q8LGD7	Q8lqd7 arabidopsis
35	300	20.1	307	2 Q9LDM2	Q9ldm2 arabidopsis
36	296	19.9	438	2 Q9M209	Q9m209 arabidopsis
37	289.5	19.4	431	2 Q6GQJ3	Q6gqj3 xenopus lae
38	283.5	19.0	291	1 RR45_SCHPO	O74918 schizosacch
39	282.5	19.0	305	1 RR45_YEAST	Q05636 saccharomyc
40	282	18.9	438	1 RR45_MOUSE	Q9jhi7 m exosome c
41	275.5	18.5	439	2 Q86Y48	Q86y48 homo sapien
42	275.5	18.5	440	2 Q6B726	Q6b726 gallus gall
43	275.5	18.5	456	2 Q86Y41	Q86y41 homo sapien
44	273.5	18.4	305	2 Q75F52	Q75f52 ashbya goss
45	272.5	18.3	286	2 Q9FP10	Q9fp10 arabidopsis

ALIGNMENTS

RESULT 1
RR43 MOUSE
ID RR43 MOUSE STANDARD; PRT; 276 AA.
AC Q9D753;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 05-JUL-2004 (Rel. 44, Last annotation update)
DE Exosome complex exonuclease RRP43 (EC 3.1.13.-) (Ribosomal RNA
processing protein 43) (Exosome component 8).
DE Names:Exosc8; Synonyms:Rrp43;
GN Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Tongue;
RX MEDLINE=22354683; PubMed=12466851; DOI=10.1038/nature01266;
RA Okazaki Y., Furuno M., Kasukawa T., Adachi J., Bono H., Kondo S.,
RA Nikaido I., Oeato N., Saito R., Suzuki H., Yamanaka I., Kiyosawa H.,
RA Yagi K., Tomaru Y., Hasegawa Y., Nogami A., Schonbach C., Gojohori T.,
RA Baldarelli R., Hill D.P., Bult C., Hume D.A., Quackenbush J.,
RA Schriml L.M., Kanapin A., Matsuda H., Batalov S., Beisel K.W.,
RA Blake J.A., Bradt D., Brusci V., Chothia C., Corbani L.E., Cousins S.,
RA Dalla E., Dragani T.A., Fletcher C.F., Forrest A., Frazer K.S.,
RA Gaasterland T., Gariboldi M., Gissi C., Godzik A., Gough J.,
RA Grimmond S., Gustincich S., Hirokawa N., Jackson I.J., Jarvis E.D.,
RA Kanai A., Kawaji H., Kawasawa Y., Kedzierski R.M., King B.D.,
RA Konagaya A., Kurochkin I.V., Lee Y., Lenhard B., Lyons P.A.,
RA Maglott D.R., Maltais L., Marchionni L., McKenzie L., Miki H.,
RA Nagashima T., Numata K., Okido T., Pavan W.J., Pertea G., Pesole G.,
RA Petrovsky N., Pillai R., Pontius J.U., Qi D., Ramachandran S.,
RA Ravasi T., Reed J.C., Reed D.J., Reid J., Ring B.Z., Ringwald M.,
RA Sadelin A., Schneider C., Sempile C.A., Setou M., Shimada K.,
RA Sultana R., Takenaka Y., Taylor M.S., Teasdale R.D., Tomita M.,
RA Verardo R., Wagner L., Wahlestedt C., Wang Y., Watanabe Y., Wells C.,
RA Wilming L.G., Wyszynski-Boris A., Yanagisawa M., Yang I., Yang L.,
RA Yuan Z., Zavolan M., Zhu Y., Zimmer A., Carninci P., Hayatsu N.,
RA Hirozane-Kishikawa T., Konno H., Nakamura M., Sakazume N., Sato K.,
RA Shiraki T., Waki K., Kawai J., Aizawa K., Arakawa T., Fukuda S.,


```
QY 133 GCGGAGCTCTGCTAGTGAAGCTGGGAACACACAGTCATTGTGGAGTTAAAGCA 192
Db 133 GCGGAGCTCTGCTAGTGAAGCTGGGAACACACAGTCATTGTGGAGTTAAAGCA 192
QY 57 AlaenGlySerAlaIleAargAlaGlyGluasnValPheValCysGlyIleAla 76
Db 57 AlaenGlySerAlaIleAargAlaGlyGluasnValPheValCysGlyIleAla 76
QY 193 GAATTTGAGCAGCAGCAGTATAGTCCCTGATAGAGGATATGTCCTTAATGTGGAC 252
Db 193 GAATTTGAGCAGCAGCAGTATAGTCCCTGATAGAGGATATGTCCTTAATGTGGAC 252
QY 77 GluileAlaGluProPheGluasnSerProasnGluGlyTrpIleValProasnLeuGlu 96
Db 77 GluileAlaGluProPheGluasnSerProasnGluGlyTrpIleValProasnLeuGlu 96
QY 253 CTACACCGCTGCTTATCATCGAGCTTCGGAGTTCGAGTCTCTGAGAGAGGCTCAAGTA 312
Db 253 CTACACCGCTGCTTATCATCGAGTTCGAGTCTCTGAGAGAGGCTCAAGTA 312
QY 97 LeuSerProLeuGluHisGlnThrLeuGlnGlnSerAsnLeuIleAsnLeuGlnSerLeu 116
Db 97 LeuSerProLeuGluHisGlnThrLeuGlnGlnSerAsnLeuIleAsnLeuGlnSerLeu 116
QY 313 ACCAGCCAGTTCATTGACAGATGCTATTGACAGAACTCACACATAATTAAGAAAGGACTTA 372
Db 313 ACCAGCCAGTTCATTGACAGATGCTATTGACAGAACTCACACATAATTAAGAAAGGACTTA 372
QY 117 ValSerGlnGluLeuHisGlnThrLeuGlnGlnSerAsnLeuIleAsnLeuGlnSerLeu 136
Db 117 ValSerGlnGluLeuHisGlnThrLeuGlnGlnSerAsnLeuIleAsnLeuGlnSerLeu 136
QY 373 TGCATTTCTCCAGGAAGCTTGTGGTCTTATCTGATCTGTCGACCTTATTGCTAGACTAC 432
Db 373 TGCATTTCTCCAGGAAGCTTGTGGTCTTATCTGATCTGTCGACCTTATTGCTAGACTAC 432
QY 137 CysilePheGluLysLysAlaAlaTrpValLeuTyrAlaAspIleIleCysLeuAsnTyr 156
Db 137 CysilePheGluLysLysAlaAlaTrpValLeuTyrAlaAspIleIleCysLeuAsnTyr 156
QY 433 GATGGGAACATTTGGATGCTGCACATTTGCTTTGTAGCAGCTTTAAAGAAAGTACAG 492
Db 433 GATGGGAACATTTGGATGCTGCACATTTGCTTTGTAGCAGCTTTAAAGAAAGTACAG 492
QY 157 AspGlySerAlaPheAspTyrAlaTrpAlaAlaLeuPheAlaLeuLysThrValLys 176
Db 157 AspGlySerAlaPheAspTyrAlaTrpAlaAlaLeuPheAlaLeuLysThrValLys 176
QY 493 TTGCTCAAGTTACTATAAATGAAGAACTGCTTTAGCGAAGTCAATTTAAAGAAAGAA 552
Db 493 TTGCTCAAGTTACTATAAATGAAGAACTGCTTTAGCGAAGTCAATTTAAAGAAAGAA 552
QY 177 LeuPro-----ThrAlaValTrpAspGluAspLeuGluArgVal 189
Db 177 LeuPro-----ThrAlaValTrpAspGluAspLeuGluArgVal 189
QY 553 AGTTATTGAATGTTAGGAAACCCAGTT-----GCTACTTCA 591
Db 553 AGTTATTGAATGTTAGGAAACCCAGTT-----GCTACTTCA 591
QY 190 IleCysAlaSerThrLeuThrArgProValGlnLeuSerThrGluValArgSerPheSer 209
Db 190 IleCysAlaSerThrLeuThrArgProValGlnLeuSerThrGluValArgSerPheSer 209
QY 592 TTGCTGTTGTTGATGACACTTCTGATGCTGATGCTGATGCTGATGCTGATGCTGATGCT 651
Db 592 TTGCTGTTGTTGATGACACTTCTGATGCTGATGCTGATGCTGATGCTGATGCTGATGCT 651
QY 210 TrpSerValPheAspAspLysLeuLeu---AlaAspProThrAspGluGluGluAspLe 228
Db 210 TrpSerValPheAspAspLysLeuLeu---AlaAspProThrAspGluGluGluAspLe 228
QY 652 GTCCACAGGAACCTTAACCGTAGTAATGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 711
Db 652 GTCCACAGGAACCTTAACCGTAGTAATGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 711
QY 228 uSerThrGluPheLeuThrIleMetLeuAsnSerSerLysAsnIleValLysIleLeuLys 248
Db 228 uSerThrGluPheLeuThrIleMetLeuAsnSerSerLysAsnIleValLysIleLeuLys 248
QY 712 GCCAGTGGGAGTGGGCTGGAGCTA 739
Db 712 GCCAGTGGGAGTGGGCTGGAGCTA 739
QY 248 sLeuGlyGlyThrHisIleGlnProLeu 257
Db 248 sLeuGlyGlyThrHisIleGlnProLeu 257
RESULT 2
D96625
hypothetical protein T2K10.14 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 09-Jul-2004
C:Accession: D96625
R:Theologos, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziali,
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
ker, M.; Wu, D.; Xu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: A86141; MUID:21016719; PMID:11130712
A:Accession: D96625
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-302 <STO>
A:Cross-references: UNIPROT:Q9ZUI4; GB:AE005173; NID:G4249388; PIDN:ADL4485.1; GSPDB:GN
C:Genetics:
A:Gene: T2K10.14
A:Map position: 1
C:Superfamily: conserved hypothetical protein MTH682
Alignment Scores:
```

```
Pred. No.: 8,61e-36 Length: 302
Score: 463.00 Matches: 99
Percent Similarity: 56.18% Conservative: 60
Best Local Similarity: 34.98% Mismatches: 105
Query Match: 31.07% Indels: 20
DB: 2 Gaps: 3
US-10-736-892-13 (1-828) x D96625 (1-302)
QY 13 TTCAAACTGTGNAACCGCTGGAGTATTACAGGAGATTTCTGAAGAAACTGCGCTCCA 72
Db 13 TTCAAACTGTGNAACCGCTGGAGTATTACAGGAGATTTCTGAAGAAACTGCGCTCCA 72
QY 19 PheAargAlilePheProLeuArgPhePheGluArgHisLeuSerGluSerLeuArgPro 38
Db 19 PheAargAlilePheProLeuArgPhePheGluArgHisLeuSerGluSerLeuArgPro 38
QY 73 GATGGAAGAGACATTGCTGAATTCACAAACACACCACTCTCAACATAGCTTCCATCAGTACA 132
Db 73 GATGGAAGAGACATTGCTGAATTCACAAACACACCACTCTCAACATAGCTTCCATCAGTACA 132
QY 39 AspGlyArgGlnLeuGlyLysAlaArgAspThrIleValAsnLeuGlyLeuValSerThr 58
Db 39 AspGlyArgGlnLeuGlyLysAlaArgAspThrIleValAsnLeuGlyLeuValSerThr 58
QY 133 GCGATGGCTGCTGCTAGTGAAGCTGGGGAACACCACTGATTTGTTGGAGTTAAAGCA 192
Db 133 GCGATGGCTGCTGCTAGTGAAGCTGGGGAACACCACTGATTTGTTGGAGTTAAAGCA 192
QY 59 AlaAspGlySerAlaLeuAlaLysIleGlySerThrThrMetLeuAlaAlaIleArgMet 78
Db 59 AlaAspGlySerAlaLeuAlaLysIleGlySerThrThrMetLeuAlaAlaIleArgMet 78
QY 193 GAATTTCCACACACACAGTAGATGCCCTGATAGAGGATATGTCGTCCTTAATGTGGAC 252
Db 193 GAATTTCCACACACACAGTAGATGCCCTGATAGAGGATATGTCGTCCTTAATGTGGAC 252
QY 79 GluValMetThrProSerThrAspSerProAspGlnGlyCysIleAlaIleGluPheHis 98
Db 79 GluValMetThrProSerThrAspSerProAspGlnGlyCysIleAlaIleGluPheHis 98
QY 253 CTACCCAGCTGTGTTTCATCGAGTTCGAGCTTCGAGCTGCTGAGAGAGGCTCAAGTA 312
Db 253 CTACCCAGCTGTGTTTCATCGAGTTCGAGCTTCGAGCTGCTGAGAGAGGCTCAAGTA 312
QY 99 MetProProIleCysSerProThrValArgProGlyArgProAlaGluAlaProVal 118
Db 99 MetProProIleCysSerProThrValArgProGlyArgProAlaGluAlaProVal 118
QY 313 ACCAGCAGTTCATTGATGATGTCATGAGAACTCACACATAATTAAGAAAGGAGACTTA 372
Db 313 ACCAGCAGTTCATTGATGATGTCATGAGAACTCACACATAATTAAGAAAGGAGACTTA 372
QY 119 IleSerLysArgLeuSerAspThrIleLeuSerSerGlyMetIleAspLeuLysGluLeu 138
Db 119 IleSerLysArgLeuSerAspThrIleLeuSerSerGlyMetIleAspLeuLysGluLeu 138
QY 373 TGCAATTTCTCAGGAGAGCTTGGTGGTTCATACGTGACCTTATTTGCTAGACTAC 432
Db 373 TGCAATTTCTCAGGAGAGCTTGGTGGTTCATACGTGACCTTATTTGCTAGACTAC 432
QY 139 CysLeuValSerGlyLysAlaAlaTrpMetGlyTyrLeuAspIleTyrCysLeuAspAla 158
Db 139 CysLeuValSerGlyLysAlaAlaTrpMetGlyTyrLeuAspIleTyrCysLeuAspAla 158
QY 433 GATGGGAACATTTGGATGCTGCACATTTGTTGTTAGCAGCTTTAAAGAAAGTACAG 492
Db 433 GATGGGAACATTTGGATGCTGCACATTTGTTGTTAGCAGCTTTAAAGAAAGTACAG 492
QY 159 AspGlyAlaLeuPheAspAlaAlaLeuAlaAlaValAlaAlaPheSerAsnLeuGln 178
Db 159 AspGlyAlaLeuPheAspAlaAlaLeuAlaAlaValAlaAlaPheSerAsnLeuGln 178
QY 493 TTGCTCAAGTTACTATAAATGAAGAACTGCTTTAGCGAAGTCAAT----- 540
Db 493 TTGCTCAAGTTACTATAAATGAAGAACTGCTTTAGCGAAGTCAAT----- 540
QY 179 IleProIleValAlaLeuAsnAspAsnGlyArgIleValAlaValThrGlyGluLysAsp 198
Db 179 IleProIleValAlaLeuAsnAspAsnGlyArgIleValAlaValThrGlyGluLysAsp 198
QY 541 -----TTAAGAGAGAAAGTATTTCGAAT 564
Db 541 -----TTAAGAGAGAAAGTATTTCGAAT 564
QY 199 GluAspAsnAlaLeuIleThrGluLysGluAlaValAsnLysGluLysArgLysLeuThr 218
Db 199 GluAspAsnAlaLeuIleThrGluLysGluAlaValAsnLysGluLysArgLysLeuThr 218
QY 565 GTTAGAGCAAAACCCAGTTGCTACTTTCATTTGCTGTGTTTGTGATGACACTTTGCTGATAGTC 624
Db 565 GTTAGAGCAAAACCCAGTTGCTACTTTCATTTGCTGTGTTTGTGATGACACTTTGCTGATAGTC 624
QY 219 LeuLysAsnIleProPheSerLeuThr---CysIleLeuHisLysAsnTyrIleLeuAla 237
Db 219 LeuLysAsnIleProPheSerLeuThr---CysIleLeuHisLysAsnTyrIleLeuAla 237
QY 625 GATCTTACCGGGAGGAGGGGACCCCTGTCACAGGAACCTTAACCGTAGTAAATGGACGA 684
Db 625 GATCTTACCGGGAGGAGGGGACCCCTGTCACAGGAACCTTAACCGTAGTAAATGGACGA 684
QY 238 AspProThrThrGluGluGluSerIleMetAspThrLeuValThrValValLeuAspSe 257
Db 238 AspProThrThrGluGluGluSerIleMetAspThrLeuValThrValValLeuAspSe 257
QY 685 GGAAGGCAAGCTGTGCTGCTCTTCCACAGCCAGGTGGGAGTGGGTGCT---GGAGGTAA 740
Db 685 GGAAGGCAAGCTGTGCTGCTCTTCCACAGCCAGGTGGGAGTGGGTGCT---GGAGGTAA 740
QY 257 rSerAspGlnMetValSerPheTyrLysSerGlyGlyAlaAlaLeuAlaTyrSerAl 277
Db 257 rSerAspGlnMetValSerPheTyrLysSerGlyGlyAlaAlaLeuAlaTyrSerAl 277
QY 741 ACTTCAGAGCTGATGATGCTGAGTCAAGTAAACAGACACAAAGAGTGTAGCACTACTCGGA 800
Db 741 ACTTCAGAGCTGATGATGCTGAGTCAAGTAAACAGACACAAAGAGTGTAGCACTACTCGGA 800
QY 277 alleLysSerCysValGluLeuAlaArgLysArgAlaLysGluLeuLysGlnIleLeuGl 297
Db 277 alleLysSerCysValGluLeuAlaArgLysArgAlaLysGluLeuLysGlnIleLeuGl 297
QY 801 TGAAGTA 807
Db 297 yGluMet 299
RESULT 3
D71032
Probable autoantigen like protein - Pyrococcus horikoshii
C:Species: Pyrococcus horikoshii
```

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: April 23, 2005, 11:46:52 ; Search time 79.5 Seconds
(without alignments)
6932.101 Million cell updates/sec

Title: us-10-736-892-13
Perfect score: 1490
Sequence: 1 atggcgccgggttcaaac.....ttcagagcatgaacacaaa 828

Scoring table: BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 1424015 seqs, 332791073 residues

Total number of hits satisfying chosen parameters: 2848030

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+n2p.model -DEV=xlp
-Q=/cgn2_1/USPTO.spool_p/US10736892/runat_22042005_114833_22551/app.query.fasta_1.967
-DB=Published Applications_AA -QFMT=fastan -SURF1=rapb -WINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOCLALIGN=200 -THR SCORE=ext -THR MAX=100
-THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=200000000 -USERR=US10736892 @CGN 1 1 130 @runat_22042005_114833_22551
-NCPUS=6 -ICPU=3 -NO MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100
-LONGLOG -DEV TIMEOUT=120 -WARN TIMEOUT=30 -THRAUS=1 -XGAPOP=10 -XGAPEXT=0.5
-FCAPOP=6 -FCAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

- 1: /cgn2_6/prodata/1/pubpaa/US07_PUBCOMB.pep.*
- 2: /cgn2_6/prodata/1/pubpaa/PCT_NEW_PUB.pep.*
- 3: /cgn2_6/prodata/1/pubpaa/US06_PUBCOMB.pep.*
- 4: /cgn2_6/prodata/1/pubpaa/US06_PUBCOMB.pep.*
- 5: /cgn2_6/prodata/1/pubpaa/US07_NEW_PUB.pep.*
- 6: /cgn2_6/prodata/1/pubpaa/PCTUS_PUBCOMB.pep.*
- 7: /cgn2_6/prodata/1/pubpaa/US08_NEW_PUB.pep.*
- 8: /cgn2_6/prodata/1/pubpaa/US08_PUBCOMB.pep.*
- 9: /cgn2_6/prodata/1/pubpaa/US09A_PUBCOMB.pep.*
- 10: /cgn2_6/prodata/1/pubpaa/US09B_PUBCOMB.pep.*
- 11: /cgn2_6/prodata/1/pubpaa/US09C_PUBCOMB.pep.*
- 12: /cgn2_6/prodata/1/pubpaa/US09_NEW_PUB.pep.*
- 13: /cgn2_6/prodata/1/pubpaa/US10A_PUBCOMB.pep.*
- 14: /cgn2_6/prodata/1/pubpaa/US10B_PUBCOMB.pep.*
- 15: /cgn2_6/prodata/1/pubpaa/US10C_PUBCOMB.pep.*
- 16: /cgn2_6/prodata/1/pubpaa/US10D_PUBCOMB.pep.*
- 17: /cgn2_6/prodata/1/pubpaa/US10_NEW_PUB.pep.*
- 18: /cgn2_6/prodata/1/pubpaa/US11_NEW_PUB.pep.*
- 19: /cgn2_6/prodata/1/pubpaa/US60_NEW_PUB.pep.*
- 20: /cgn2_6/prodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result Query

No.	Score	Match	Length	DB	ID	Description
1	466	31.3	302	15	US-10-424-599-277230	Sequence 277230, Sequence 277231, Sequence 63522, A
2	456	30.6	307	15	US-10-424-599-277231	Sequence 63522, A
3	415	27.9	347	15	US-10-425-114-63522	Sequence 65625, A
4	413.5	27.8	346	15	US-10-425-114-65625	Sequence 65625, A
5	406	27.2	275	16	US-10-437-963-119834	Sequence 119834, Sequence 47476, A
6	394	26.4	334	15	US-10-425-114-47476	Sequence 47476, A
7	300	20.1	430	15	US-10-425-114-64425	Sequence 64425, A
8	299.5	20.1	396	16	US-10-437-963-144343	Sequence 144343, Sequence 13, Appl
9	282	18.9	438	16	US-10-451-861-13	Sequence 625, Appl
10	260.5	17.5	287	15	US-10-310-154-625	Sequence 38982, A
11	255.5	17.1	207	16	US-10-767-701-38982	Sequence 38982, A
12	236	15.8	108	15	US-10-108-260A-3952	Sequence 3952, Ap
13	232.5	15.6	323	15	US-10-424-599-254036	Sequence 254036, Sequence 254030, A
14	227.5	15.3	286	15	US-10-424-599-254030	Sequence 254030, A
15	202	13.6	124	16	US-10-767-701-62295	Sequence 62295, A
16	198	13.3	337	14	US-10-032-585-7279	Sequence 7279, Ap
17	176.5	11.8	154	15	US-10-424-599-261519	Sequence 261519, Sequence 39155, A
18	143	9.6	192	15	US-10-425-114-39155	Sequence 39155, A
19	139.5	9.4	123	16	US-10-767-701-57278	Sequence 57278, A
20	134.5	9.0	359	15	US-10-282-122A-45707	Sequence 45707, A
21	134	9.0	168	16	US-10-437-963-129437	Sequence 129437, Sequence 7080, Ap
22	131	8.8	350	14	US-10-032-585-7080	Sequence 664, App
23	124.5	8.4	256	9	US-09-925-298-664	Sequence 664, App
24	124.5	8.4	256	14	US-10-102-806-664	Sequence 664, App
25	122.5	8.2	361	15	US-10-282-122A-57735	Sequence 57735, A
26	117	7.9	245	10	US-09-893-519A-48	Sequence 48, Appl
27	116	7.8	278	15	US-10-425-114-62410	Sequence 62410, A
28	116	7.8	292	15	US-10-425-114-49949	Sequence 49949, A
29	115	7.7	223	16	US-10-767-701-32772	Sequence 32772, A
30	115	7.7	260	15	US-10-425-114-61823	Sequence 61823, A
31	115	7.7	271	15	US-10-425-114-68490	Sequence 68490, A
32	113.5	7.6	245	15	US-10-258-662-1	Sequence 1, Appl
33	113	7.6	242	16	US-10-437-963-121395	Sequence 121395, Sequence 260580, Sequence 227970, A
34	112.5	7.6	241	15	US-10-424-599-260580	Sequence 260580, Sequence 227970, A
35	109	7.3	254	15	US-10-424-599-227970	Sequence 2044, Ap
36	108	7.2	191	15	US-10-264-237-2044	Sequence 173123, Sequence 42, Appl
37	107	7.2	259	15	US-10-424-599-173123	Sequence 42, Appl
38	105	7.0	705	10	US-09-907-907A-44	Sequence 44, Appl
39	105	7.0	705	10	US-09-907-907A-42	Sequence 45571, A
40	104	7.0	362	15	US-10-282-122A-45571	Sequence 6245, Ap
41	102.5	6.9	245	9	US-09-738-626-6245	Sequence 168217, Sequence 2004, Ap
42	101.5	6.8	118	16	US-10-437-963-168217	Sequence 10721, A
43	101.5	7.1	725	15	US-10-369-493-2004	Sequence 57142, A
44	101	6.8	451	9	US-09-815-242-10721	
45	100	6.7	451	15	US-10-282-122A-57142	

ALIGNMENTS

RESULT 1
US-10-424-599-277230
; Sequence 277230, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 277230
; LENGTH: 302
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_92360C.1.pep
; US-10-424-599-277230

Alignment Scores:

Pred. No.: 1,21e-40 Length: 302
 Score: 466.00 Matches: 99
 Percent Similarity: 59.15% Conservative: 69
 Best Local Similarity: 34.86% Mismatches: 101
 Query Match: 31.28% Indels: 16
 DB: 15 Gaps: 3

US-10-736-892-13 (1-828) x US-10-424-599-277230 (1-302)

```

Qy 13 TCAAACTGGGAACCGCTGAGTATTACAGGAGATTCTGAAAGAAAACCTGCCGTCCA 72
Db   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
Qy 19 PheArgLeuPheProLeuArgTyrPheGluArgHisLeuAlaGluSerIleArgPro 38
Db   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
Qy 73 GATGGAGAGACATTGGTGAATTGAGAACCACTGTCAACATAGTTCCATGAGTACA 132
Db   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
Qy 39 AspGlyArgProLeuGlyLeuAlaArgGluThrSerIlePheLeuGlyAlaValAlaSer 58
Db   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
Qy 133 CGGATGCTCTGCTAGTGAAGCTGGGGAACACCACTGTCATTGTGGAGTTAAAGCA 192
Db   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
Qy 59 AlaGlySerAlaLeuValIleGlySerThrIleLeuThrAlaIleGlyMet 78
Db   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
Qy 193 GAATTGTCAGCACCACTGATGCCCTGATAGAGATATGTCCTCCCTAAATGTGGAC 252
Db   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
Qy 79 GluValMetThrProSerLeuGluSerProAspGluGlyCysLeuAlaIleAspPheHis 98
Db   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
Qy 253 CTACCAACCGCTGTTCTATCGAGGTTTCGGACTGACCTCTGGAGAGAGGCTCAAGTA 312
Db   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
Qy 99 MetProIleCysSerProIleValArgProGlyArgProAlaGluAlaSerProVal 118
Db   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
Qy 313 ACCAGCAGTTCATTGACAGATGTCATTGAGAACTCACACATAATTAAGAAAGAGACTTA 372
Db   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
Qy 119 ValSerLysGlnLeuSerAspThrIleSerSerLysMetIleAspLeuLysGluLeu 138
Db   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
Qy 373 TGCATTTCTCAGGGAACCTGTTGGTTCTATCTGTGACCTTATTTGCTAGACTAC 432
Db   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
Qy 139 SerLeuValSerGlyLeuAlaAlaTrpMetAlaTyrLeuAspIleTyrCysLeuAspAla 158
Db   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
Qy 433 GATGGGAACATTGTCAGCTGTCACATTTCTGTTGTAGCAGCTTTAAAGAAAGTACAG 492
Db   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
Qy 159 AspGlyAlaLeuPheAspAlaLeuLeuSerAlaValAlaSerLeuSerHisLeuGln 178
Db   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
Qy 493 TTGCTCTGAAGTTACTATAAATCAAGAAACTGCTTTAGCGGAAGTCAAT----- 540
Db   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
Qy 179 IleProValValAlaMetAsnAspAspGlyLysIleValLeuValSerGluLysGly 198
Db   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
Qy 541 -----TTAAGAAAGAAAGTTATTGTAATGTTTAGAGCAAC 576
Db   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
Qy 199 GlnLysArgGluGlnProValAsnLysGlnLysArgLysLeuThrLeuArgSerIle 218
Db   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
Qy 577 CCAGTTGCTACTCATTTGCTGTTGATGACACTTTGCTGATGTCGATCTTACCGGG 636
Db   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
Qy 219 ProPheSerLeuThr---CysIleLeuHisLysAsnTyrIleLeuAlaAspProThrAla 237
Db   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
Qy 637 GAGGAGGGACCTGTCTCCACAGGAACCTTAACCGTAGTAAATGACAGCAAGCAAGCT 696
Db   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
Qy 238 GluGlu-GluSerIleMetGluThrHisValThrValValLeuAspThrSerGlyGlnLe 257
Db   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
Qy 697 GTGCTGTCTTCAAGCCAGCTGGAGTGG-----GCTGCTGGAGCTAACTTCAGGACTG 752
Db   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
Qy 257 uileSerLeuTyrLysProGlyGlyProValLeuAlaTyrThrSerAlaIleGlnAspCy 277
Db   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
Qy 753 CATGAGTCGACAGTAACAGACACAAAGAGTGGCAAACTACTGGATGAAGTAATTC 812
Db   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
Qy 277 sAlaAlaLeuThrArgGlnArgValLysGluLeuLysSerPheLeuAspLysAlaAsnSe 297
Db   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
Qy 813 GAGCATGAAA 822
Db   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
Qy 297 rAlaMetGlu 300
Db   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::

```

RESULT 2

US-10-424-599-277231

; Sequence 277231, Application US/10424599
 ; Publication No. US20040031072A1
 ; GENERAL INFORMATION:
 ; APPLICANT: La Rosa Thomas J
 ; APPLICANT: Kovalic David K
 ; APPLICANT: Zhou Yihua
 ; APPLICANT: Cao Yongwei
 ; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
 ; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
 ; FILE REFERENCE: 38-21(53223)B
 ; CURRENT APPLICATION NUMBER: US/10/424,599
 ; CURRENT FILING DATE: 2003-04-28
 ; NUMBER OF SEQ ID NOS: 285684
 ; SEQ ID NO 277231
 ; LENGTH: 307
 ; TYPE: PRT
 ; ORGANISM: Glycine max
 ; FEATURE:
 ; OTHER INFORMATION: Clone ID: PAT_MRT3847_92361C.1.pep
 ; US-10-424-599-277231

Alignment Scores:
 Pred. No.: 1,43e-39 Length: 307
 Score: 456.00 Matches: 99
 Percent Similarity: 58.33% Conservative: 69
 Best Local Similarity: 34.38% Mismatches: 101
 Query Match: 30.60% Indels: 20
 DB: 15 Gaps: 4

US-10-736-892-13 (1-828) x US-10-424-599-277231 (1-307)

```

Qy 13 TCAAACTGGGAACCGCTGAGTATTACAGGAGATTCTGAAAGAAAACCTGCCGTCCA 72
Db   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
Qy 19 PheArgLeuPheProLeuArgTyrPheGluArgHisLeuAlaGluSerIleArgPro 38
Db   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
Qy 73 GATGGAGAGACATTGGTGAATTGAGAACCACTGTCAACATAGTTCCATGAGTACA 132
Db   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
Qy 39 AspGlyArgProLeuGlyLysAlaArgGluThrSerIlePheLeuGlyAlaValAlaSer 58
Db   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
Qy 133 CGGATGCTCTGCTAGTGAAGCTGGGGAACCACTGTCATTGTGGAGTTAAAGCA 192
Db   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
Qy 59 AlaGlySerAlaLeuValIleGlySerThrThrMetLeuThrAlaIleLysMet 78
Db   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
Qy 193 GAATTTCTCAGCACCACTGATGCTCCCTGATAGAGATATGTCGCTCCCTAAATGTGGAC 252
Db   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
Qy 79 GluValMetThrProSerLeuGluSerProAspGluGlyCysLeuAlaIleAspPheHis 98
Db   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
Qy 253 CTACCAACCGCTGTTCTATCGAGTTCGAGCTGGACCTCTCTGGAGAGAGGCTCAAGTA 312
Db   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
Qy 99 MetProIleCysSerProIleValArgProGlyArgProAlaGluAlaSerProVal 118
Db   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
Qy 313 ACCAGCAGTTCATTGACAGATGTCATTGAGAACTCACACATAATTAAGAAAGAGACTTA 372
Db   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
Qy 119 ValSerLysGlnLeuSerAspThrIleSerSerArgMetIleAspLeuLysGluLeu 138
Db   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
Qy 373 TGCATTTCTCAGGGAACCTGTTGGTTCTATCTGTGACCTTATTTGCTAGACTAC 411
Db   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
Qy 139 SerLeuValGlyGlyLysAlaAlaTrpMetAlaTyrLeuAsnAlaAlaPheSerIleMet 158
Db   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
Qy 412 ---GACCTTATTGCTTAGACTACGATGGGAACATTTTGGATGCTGCACATTTGCTTTG 468
Db   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
Qy 159 GlnAspIleTyrCysLeuAspAlaAspGlyAlaLeuPheAspAlaAlaLeuSerAla 178
Db   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
Qy 469 TTAGCAGCTTTAAAGATGTACAGTTCCTGCTGAGTACTATAATTAAGAAAGAGCTTTA 528
Db   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
Qy 179 ValAlaAlaLeuSerHisLeuGlnIleProAlaAlaMetAsnAspGlyLysIle 198
Db   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
Qy 529 GCGGAAGTCAAT-----TTAAGAAAGAAAGTTATTTTGAAT 564
Db   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
Qy 199 ValLeuValSerGlyGluAspGlyGlnLysProValAsnLysGluLysArgLysLeuThr 218
Db   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
Qy 565 GTTAGACAAACCCAGTTGCTTACTTCTGTTGTTGATGACACTTTGCTGATAGTGC 624
Db   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::

```


OTHER INFORMATION: Incyte ID No. 6673549 2551987CD1

Db ValGlnGluAspLeuArgValAspGlyArgGlyCysGluAspTyrArgCysValGluVal 37
QY 112 AACATAGTTTCATCAGTACAGCGATCGCTCTCTAGTGAAGCTGGGAAACACACACA 171
Db GluThrAspValValSerAsnThrSerGlySerAlaArgValLysLeuGlyHisThrAsp 57
QY 172 GTCATTTTGGAGTTAAAGCAGAAATTTCCAGCACCACCATGAGTGCCTGATAGAGGA 231
Db IleuValGlyValLysAlaGluMetGlyThrProLysLeuGluLysProAsnGluGly 77
QY 232 TATGTCGCTTAATGTGGACCTACACCGCTGTGTTTCATCGAGGTTTCGGAGTGGACCT 291
Db TyrLeuGluPhePheValAspCysSerAlaSerAlaThrProGluPheGlu---GlyArg 96
QY 292 CCTGGAGAGAG---GCTCAAGTAAACAGCAGCTTCATTCGAGATGTCATTCAGAACTCA 348
Db GlyGlyAspAspLeuGlyThrGluLeuAlaAsnThrLeuTyrArgIlePheAsnAsnLys 116
QY 349 CACATAATTAAGAAAGAGGACTTATGCAATTTCTCCAGGAAAGCTTGTCTGGTTCTATAC 408
Db SerSerValAspLeuLysThrLeuCysIleSerProArgGluHisCysTyrValLeuTyr 136
QY 409 TGTGACCTTATTCGCTAGATACAGATGGGAACATTTTGGATGCTGCACATTTGCTTTG 468
Db ValAspValLeuLeuLeuGluCysGlyGlyAsnLeuPheAspAlaIleSerIleAlaVal 156
QY 469 TTACGAGCTTTAAAGAAATGACAGTTCCTCAAGTT---ACTATAATGAAGAACTGCT 525
Db LysAlaAlaLeuPheAsnThrArgIleProArgValArgValLeuGluAspGluGluGly 176
QY 526 TTAGCGGAAGTCAATTTAAAGAAAGAAAGTTAT-----TTGAATGTTAGAGCA 573
Db SerLysAspIleGluLeuSerAspAspProTyrAspCysIleArgLeuSerVal-Glu-- 195
QY 574 AACCAGTTGCTACTTCATTTGCTGTGTTGATGACACTTTGCTG----- 618
Db -----AsnValProCysIleValThrLeuCysLysIleGlyTyrArgHi 210
QY 619 -ATAGTCGATCCTACCGGGAGGAGGGCCACCTGTCACAGAACTTACCTAGTAA 677
Db sValValAspAlaThrLeuGlnGlu-GluAlaCysSerLeuAlaSerLeuValSerV 230
QY 678 TGGACGAGAGGCAAGCTGCTCTCTTCACAAAGCCAGTGGGAGTGGCTGCT-GGAG 736
Db alThrSerLysGlyValValThrCysMetArgLysValGlyLysGlySerLeuAspProG 250
QY 737 CTAAACTTCAGACTGCTAGTGCAGCAGTAAAGAGACACAAAGAGTGAAGCACTAC 796
Db LuSerIlePheGluMetMetGluThrGly-----LysArgValGlyLysValL 266
QY 797 TGGATGAAGTAAATTCAGAGCATGAACACAAA 828
Db euHisAlaSerLeuGlnSerValLeuHisLys 276

RESULT 2

US-09-248-796A-17015
; Sequence 17015, Application US/09248796A
; Patent No. 6747137
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstock et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN
; TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.132
; CURRENT APPLICATION NUMBER: US/09/248,796A
; CURRENT FILING DATE: 1999-02-12
; PRIOR APPLICATION NUMBER: US 60/074,725
; PRIOR FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: US 60/096,409
; PRIOR FILING DATE: 1998-08-13
; NUMBER OF SEQ ID NOS: 28208
; SEQ ID NO 17015
; LENGTH: 349
; TYPE: PRT

; ORGANISM: Candida albicans
US-09-248-796A-17015
Alignment Scores:
Pred. No.: 5,94e-14 Length: 349
Score: 192.00 Matches: 67
Percent Similarity: 40.14% Conservative: 45
Best Local Similarity: 24.01% Mismatches: 110
Query Match: 12.89% Indels: 57
DB: 4 Gaps: 7

US-10-736-892-13 (1-828) x US-09-248-796A-17015 (1-349)

QY 37 TATTACAGGAGATTCTTGAAGAAATGCGCGT-----CCAGAT 75
Db 30 TyrLeuIleGlnAlaLeuLysSerAsnIleArgLeuSerThrSerThrAsnThrGlyGly 49
QY 76 GGAAGAGAACTTGTGTAATTCAGAACCCACAACTGTCAACATAGTTCGATCAGTACAGCG 135
Db 50 SerArgLysPheAsnGlnPheArgProIleAspIleLysLeuSer-----AsnThrArg 67
QY 136 GATGGCTCTCTCTAGTGAAGCTGGGGAACACACAGTCTATTGTGGAGTTAAAGCAGAA 195
Db 68 TyrGlySerValGluLeuSerLeuGlyLysThrLysValMetValAsnIleThrSerArg 87
QY 196 TTTGCAGCACCACTAGATGCCCTGATAGAGATATGCTGCTCCCTAATGTGGACCTA 255
Db 88 IleThrGluProTyrGlnAspArgProPheGluGlyIleMetThrIleAsnCysGluIle 107
QY 256 CCA-----CGCTGTGTTTCATCGAGGTTTCGGAGTTCGAGCTCTCTGGAGAGAGGCT 306
Db 108 ProAsnHisIleLysLeuGlnSerAsnAspThrThrAsnAsnAsnAsnAspAsnGluAsp 127
QY 307 CAA-----GTAACCCAGCTTCATTCGAGATGCTCATTGAGAACTCAGACATATT 357
Db 128 GluPheIleAsnLeuIleAsnArgAlaLeuAspArgAlaIleArgArgSerAsnAlaVal 147
QY 358 AAGAAGAGACATTATGCTATTTCTCCAGGGAAGCTTCTGTTGGTTCTATCTGTGACCTT 417
Db 148 AspLeuGlnAsnLeuCysIleIleAlaGlyGlnLysValTrpGluLeuIleIleAspLeu 167
QY 418 ATTTGCTAGACTACGATGGGAACATTTTGGATGCCCTGACATTTGCTTTGTAGCAGCT 477
Db 168 GlnValLeuAsnTyrAspGlyAsnLeuIleAspSerGlyCysLeuAlaIleIleThrAla 187
QY 478 TTAAGATCTACAGTTGCTGCTGAGTACTTAAATGAAGAACTGCTTTAGCGGAAGTC 537
Db 188 LeuLeuAspPheLysLysProAspValThrIleAsnAsnAsn----- 201
QY 538 AATTTAAAGAAAGAAAGTTAT----- 558
Db 202 AsnAsnSerAsnSerSerTyrThrSerGlyGlyGlyAsnIleIleValHisAspAsp 221
QY 559 -----TTGAATGTTTGAAGCAACCCAGTGTGCTACTTCAATT 594
Db 222 GluMetLysArgProPheIleGluLeuSerIleLeuHisIleProIleCysLeuThrPhe 241
QY 595 GCTGTGTTT-----GATGACACT 612
Db 242 ValLeuPheAsnLeuGlySerLysGluThrAsnLeuLysThrAsnAspIleAspGlnGlu 261
QY 613 TTGCTGATAGTCTGATCTCCCGGGAGGAGGGCCACCTGTCACAGAAACCTTAACCGT 672
Db 262 IleTrpLeuLeuArgGlyAspAlaMetGlu-GluSerCysArgAspGlyTyrLeuIleMe 281
QY 673 AGTAATGACGAGAAAGCAAGCTGTGCTCTTCAACAGCCAGGTGGAGTGG 727
Db 281 tThrMetAsnGlnAsnHisGluLeuIleGlnLeuSerLysIleGlyGlyGly 299

RESULT 3

US-09-270-767-32918
; Sequence 32918, Application US/09270767
; Patent No. 6703491

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: April 23, 2005, 10:13:00 ; Search time 93.5 Seconds
(without alignments)
6850.002 Million cell updates/sec

Title: US-10-736-892-13

Perfect score: 1490

Sequence: 1 atggggcggtttcaaac.....ttcagagcatgaacacacaaa 828

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2105692 seqs, 386760381 residues

Total number of hits satisfying chosen parameters: 4211384

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ n2p.model -DEV=xlp
-O=/cpn2.1/USPTO spool_p/US10736892/runat_22042005_114830_22427/app.query.fasta_1.967
-DB=A_Geneseq_16Dec04 -QFMT=fastan -SUFFIX=rag -MINMATCH=0.1 -LOOPEXT=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10736892 @CGN 1.154 @runat_22042005_114830_22427 -NCPU=6 -ICPU=3
-NO WMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : A_Geneseq_16Dec04:.*
1: Geneseq1980s:.*
2: Geneseq1990s:.*
3: Geneseq2000s:.*
4: Geneseq2001s:.*
5: Geneseq2002s:.*
6: Geneseq2003as:.*
7: Geneseq2003ba:.*
8: Geneseq2004s:.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1313	88.1	276	5	Abb97415 Novel hum
2	1313	88.1	276	8	Abm81511 Tumour-as
3	1313	88.1	276	8	Adp55135 Human PRO
4	806	54.1	171	4	Aab63406 Human bre
5	806	54.1	171	4	Aab63276 Human bre
6	773	51.9	157	4	Aab63405 Human bre
7	485	32.6	113	8	Adr08498 Human pro
8	477	32.0	112	4	Aab63407 Human bre
9	464	31.1	289	3	Aag29331 Arabidops
10	464	31.1	352	3	Aag29329 Arabidops

11	463	31.1	302	3	AAG29330	Arabidops
12	369.5	24.8	272	8	ADM47715	Thermococ
13	342.5	23.0	267	7	ADM25774	Hyperther
14	300	20.1	300	3	AAG51508	Arabidops
15	300	20.1	300	3	AAG23307	Arabidops
16	300	20.1	307	3	AAG51507	Arabidops
17	300	20.1	307	3	AAG23306	Arabidops
18	282.5	19.0	305	6	ABR52363	Protein s
19	282.5	19.0	305	7	ADK62640	Disease t
20	282	18.9	438	5	ABB83339	Murine TS
21	269	18.1	291	7	ADC131596	Human nov
22	269	18.1	291	8	ADL12597	Human ste
23	269	18.1	291	8	ABM80445	Tumour-as
24	269	18.1	291	8	ADP55026	Human PRO
25	269	18.1	299	7	ADC33202	Human nov
26	264.5	17.8	423	7	ADP59459	Human PRO
27	260.5	17.5	287	8	ADM48207	Polypepti
28	236	15.8	108	7	ADM05267	Human pro
29	235.5	15.8	247	3	AAG23308	Arabidops
30	235.5	15.8	247	3	AAG51509	Arabidops
31	235.5	15.8	292	8	ADR85972	Aspergill
32	227	15.2	281	4	ABB61929	Drosophil
33	206	13.8	125	5	ABP01119	Human ORF
34	198	13.3	337	5	ABP73442	Candida a
35	189.5	12.7	355	8	ADQ14332	Human 75
36	189.5	12.7	355	8	ADQ08051	Human SPA
37	189.5	12.7	355	8	ADR41713	Protein s
38	189.5	12.7	372	8	ADQ14333	Human 75
39	189.5	12.7	372	8	ADR41714	Protein s
40	189.5	12.7	466	4	ABE64530	Drosophil
41	176	11.8	218	8	ADK16271	Nanoarcha
42	143.5	9.6	394	6	ABR52962	Protein s
43	143.5	9.6	394	7	ADK62638	Disease t
44	134.5	9.0	359	6	ABU17783	Protein e
45	131	8.8	350	5	ABP73243	Candida a

ALIGNMENTS

RESULT 1

ABB97415
ID ABB97415 standard; protein; 276 AA.

XX AC ABB97415;

XX DT 27-JUN-2002 (first entry)

XX DE Novel human protein SEQ ID NO: 683.

XX KW Human; antianaemic; vulnery; antiinflammatory; immunomodulator;
KW antinfertility; cerebroprotective; cytostatic; rheumatic; gene therapy;
KW neuroprotective; antiparkinsonian; protein therapy; EST;
XX expressed sequence tag.

XX OS Homo sapiens.

XX PN WO200222660-A2.

XX PD 21-MAR-2002.

XX PF 10-SEP-2001; 2001WO-US026015.

XX PR 11-SEP-2000; 2000US-00659671.

XX PA (HYSE-) HYSEQ INC.

XX PI Tang YT, Liu C, Zhou P, Asundi V, Zhang J, Zhao QA, Ren F;

XX PI Xue AJ, Yang Y, Zehrman T, Drmanac RT;

XX WPI: 2002-292408/33.

XX DR N-PSDB; ABB32601.

XX PT An isolated polynucleotide for treating diseases associated with its

PT encoded polypeptide such as cancer and multiple sclerosis.

PS Claim 20; SEQ ID NO 683; 509pp; English.

XX The present invention provides the protein and coding sequences of 444 novel human proteins. These were isolated from expressed sequences tags (ESTs). They can be used to stimulate cell growth, to regulate haematopoiesis e.g. to treat aplastic anaemia, to help tissue regrowth e.g. in burn treatment, to regulate the immune system e.g. to treat multiple sclerosis, to regulate activin or inhibin e.g. to treat infertility, to regulate haemostasis or thrombolysis e.g. to treat stroke and cancer, to screen for drugs, to treat inflammatory conditions e.g. rheumatoid arthritis, and to treat nervous system disorders e.g. Parkinson's disease. The present sequence is a protein of the invention

XX Sequence 276 AA;

Alignment Scores:

Pred. No.:	1.32e-143	Length:	276
Score:	1313.00	Matches:	257
Percent Similarity:	96.03%	Conservative:	9
Best Local Similarity:	92.78%	Mismatches:	10
Query Match:	88.12%	Indels:	2
DB:	5	Gaps:	0

US-10-736-892-13 (1-828) x ABB97415 (1-276)

QY	1	ATGCGCGCGGTTC	AAACTGTGGACCGCTGGAGTATTACAGGAGATTCTGAAGAA	60
DB	1	MetAlaAlaGlyPhe	LeuThrValGluProLeuGluTyrTyrArgArgPheLeuLeuGlu	20
QY	61	AACCTGCCGTCAG	ATGGAAGAACTTGGTGAATTCAGAACCACTGTCAACATAGGT	120
DB	21	AsnCysArgPro	aspGlyArgGluLeuGlyGluPheArgThrThrValAsnIleGly	40
QY	121	TCGATCAGTACAG	CGGATGCTGTCTAGTGAAGCTGGGAAACACCAAGTCATTGT	180
DB	41	SerIleSerThr	AlaAspGlySerAlaLeuValLeuGlyAsnThrThrValIleCys	60
QY	181	GGAGTTAAACGAG	CAATTTGAGACACCACTGAGTCCCTCATAGAGGATGCGTC	240
DB	61	GlyValLeuAla	GluPheAlaProSerThrAspAlaProAspGlyTyrValVal	80
QY	241	CCTAATGTGACCT	TACACACCGCTGTGTTCATCGAGGTTTCGAGTCCCTCATAGAGGATGCGTC	300
DB	81	ProAsnValAsp	LeuProLeuCysSerSerArgPheArgSerGlyProProGlyGlu	100
QY	301	GAGCTCAAGTAA	CCGCCAGTTCATTGCGAGTGTATTGAGAACTCACATTAATTAAG	360
DB	101	GluAlaGlnVal	AlaSerGlnPheIleAlaAspValIleGluAsnSerGlnIleGln	120
QY	361	AAAGAGGACTTAT	GCAATTTCTCCAGGGAAGCTTGGTGTCTATCTGCTGACCTTATT	420
DB	121	LysGluAspLeu	CysIleSerProGlyLeuValTrpValLeuTyrCysAspLeuIle	140
QY	421	TGCTAGACTACG	ATGGCAACATTTGGATGCTCTGCACTTTGTTGTAGAGCTTTA	480
DB	141	CysLeuAspTyr	AspGlyAsnIleLeuAspAlaCysThrPheAlaLeuLeuAlaLeu	160
QY	481	AGATATGACATG	TGCTGAGTACTATAAATGAAGAACTGCTTTAGCGGAAGTCAAT	540
DB	161	LysAsnValGln	LeuProGluValThrIleAsnGluLeuAlaGluValAsn	180
QY	541	TTAAAGAGAAAG	TATTATTTGAATGTTAGACCAACCCAGTGTCTACTTCTATTGCTGTG	600
DB	181	LeuLysLysSer	TyrLeuAsnIleArgThrHisProValAlaThrSerPheAlaVal	200
QY	601	TTTGATCACACT	TGTGCTAGTGCATCTCAACGGGAGGAGGGCCACCTGTGCCACAG	660
DB	201	PheAspAspThr	LeuLeuIleValAspProThrGlyGluGluHis-LeuAlaThrGln	220
QY	661	AACCTTAACGTA	GTATGACGAGGGAAGGCAAGCTGTCTCTTCAACAGGCAGGTGG	720

DB	220	YThrLeuThrIleValMetAspGluGluGlyLeuLeuCysLeuHisLysProGlyGln	240
QY	721	GAGTGGGCTGCT-GGAGCTAAACTTCAGGCTCAGTGCAGGACGAGTAAACGACACAA	779
DB	240	YSerGlyLeuThrGlyAlaLysLeuGlnAspCysMetSerArgAlaValThrArgHisLys	260
QY	780	AGAAGTGAGCAAACTACTGCTGAAGTAAATTCAGAGCATGCAACACAA	828
DB	260	SGluValLysLysLeuMetAspGluValIleLysSerMetLysProLys	276
RESULT 2			
ABM81511			
ID	ABM81511	standard; protein; 276 AA.	
XX	AC	ABM81511,	
XX	DT	18-NOV-2004 (first entry)	
XX	DE	Tumour-associated antigenic target (TAT) polypeptide PRO82291, SEQ:3894.	
XX	DE	Tumour-associated antigenic target; TAT; human; overexpression; cancer;	
XX	KW	tumour; diagnosis; cell proliferative disorder; breast cancer;	
XX	KW	colorectal cancer; lung cancer; ovarian cancer; liver cancer;	
XX	KW	central nervous system cancer; bladder cancer; pancreatic cancer;	
XX	KW	cervical cancer; melanoma; leukaemia; hybridisation probe;	
XX	KW	chromosome identification; chromosome mapping; gene mapping;	
XX	OS	gene therapy; cytostatic.	
XX	OS	Homo sapiens.	
XX	PN	WO2004030615-A2.	
XX	PD	15-APR-2004.	
XX	PF	29-SEP-2003; 2003WO-US028547.	
XX	PR	02-OCT-2002; 2002US-0414971P.	
XX	PA	(GETH) GENENTECH INC.	
XX	PI	Wu TD, Zhang Z, Zhou Y;	
XX	DR	WPI; 2004-347921/32.	
XX	DR	N-PSDB; ACN39626.	
XX	PT	New tumor-associated antigenic target polypeptides and nucleic acids,	
XX	PT	useful in preparing a medicament for treating or detecting a	
XX	PT	proliferative disorder, e.g. breast, lung, colorectal, ovarian or	
XX	PT	prostate cancer or tumor.	
XX	PS	Claim 12; SEQ ID NO 3894; 7273pp; English.	
XX	CC	The invention relates to human tumour-associated antigenic target (TAT)	
XX	CC	polypeptides, and their related nucleic acids. The TAT polypeptides are	
XX	CC	overexpressed in cancer tissues compared to normal tissues, and may thus	
XX	CC	serve as effective targets for the diagnosis and treatment of cancer in	
XX	CC	mammals. The invention also relates to nucleic acid and polypeptide	
XX	CC	sequences at least 80% identical to the TAT nucleic acids and	
XX	CC	polypeptides; expression vectors and host cells comprising a TAT nucleic	
XX	CC	acid; an antibody specific for a TAT polypeptide; a peptide or organic	
XX	CC	molecule which binds to a TAT polypeptide; fusion proteins comprising a	
XX	CC	TAT polypeptide; and methods and compositions for the treatment or	
XX	CC	diagnosis of cancer in mammals. TAT polypeptides, nucleic acids,	
XX	CC	antibodies, antagonists, binding molecules and compositions are useful	
XX	CC	for diagnosing or treating a cell proliferative disorder associated with	
XX	CC	increased TAT expression, particularly cancers such as breast cancer,	
XX	CC	colorectal cancer, lung cancer, ovarian cancer, liver cancer, bladder	
XX	CC	cancer, pancreatic cancer, cervical cancer, cancers of the central	
XX	CC	nervous system, melanoma and leukaemia. TAT nucleic acids may further be	
XX	CC	used as hybridisation probes, in chromosome and gene mapping, in	
XX	CC	chromosome identification and in gene therapy. The present sequence	
XX	CC	represents a TAT polypeptide of the invention	

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: April 23, 2005, 12:47:30 ; Search time 3101 Seconds
(without alignments)
3387.855 Million cell updates/sec

Title: US-10-736-892-12

Perfect score: 1437

Sequence: 1 MAAGFTVPLEYRRLKE.....TRHKEVSKLLDEVQSMKHK 276

Scoring table:

	BLOSUM62	Xgapop 10.0	Xgapext 0.5
Ygapop 10.0			0.5
Ygapop 6.0			7.0
Delop 6.0			7.0

Searched: 34239544 seqs, 19032134700 residues

Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

```
-MODEL=frame+ p2n.model -DEV=xlp
-O=/cgn2 1/USPTO.spool_p/US10736892/runat 22042005 115016 25332/app_query.fasta_1.455
-DB=EST -QPM=fastap -SUFFIX=rat -MINMATCH=0.1 -LOPCL=0 -LOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=200000000
-USER=US10736892 @CGN 1 1 5180 @runat 22042005 115016 25332 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7
```

Database :

```
EST:*
1: gb_est1:*
2: gb_est2:*
3: gb_est3:*
4: gb_est4:*
5: gb_est5:*
6: gb_est6:*
7: gb_est7:*
8: gb_est8:*
9: gb_est9:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1345	93.6	943	3 AK009584	Mus muscu
2	1279	89.0	1012	6 BY709262	BY709262
3	1257	87.5	1025	5 BM904094	BM904094
4	1255.5	87.4	924	4 BI853518	BI853518
5	1246.5	86.7	941	3 CR595030	CR595030
6	1239	86.2	996	6 BY762333	BY762333
7	1228	85.5	944	1 AL531894	AL531894
8	1226	85.3	935	6 CB204353	CB204353
9	1224.5	85.2	896	3 CR612849	CR612849

10	1224.5	85.2	903	3	CR602904	full-leng
11	1214	84.5	866	7	CN644054	ILLUMIGEN
12	1188.5	82.7	811	9	AY400223	Mus muscu
13	1188	82.7	820	2	BE738362	BE738362
14	1184	82.4	860	5	BQ223602	AGENCOURT
15	1171	81.5	854	4	BI603115	603252148
16	1163	80.9	1098	4	BM467670	AGENCOURT
17	1157	80.5	819	4	BI152376	602917558
18	1147	79.8	841	5	EX425474	EX425474
19	1140	79.3	877	1	AL558475	AL558475
20	1137	79.1	879	7	CN153352	940460 MA
21	1134	78.9	747	7	CN348606	170006001
22	1131	78.7	951	5	BQ224502	AGENCOURT
23	1129.5	78.6	811	9	AY400221	Homo sapi
24	1114	77.5	888	1	AL558476	AL558476
25	1114	77.5	888	5	BQ277403	AGENCOURT
26	1107.5	77.1	810	7	CF579802	AGENCOURT
27	1094	76.1	740	4	BG568455	602587419
28	1090	75.9	751	6	CB950339	AGENCOURT
29	1089	75.8	874	1	AJ819699	AJ819699
30	1088	75.7	739	7	CN155464	942764 MA
31	1084	75.4	738	6	CB311683	AGENCOURT
32	1079	75.1	721	7	CO734601	SI1H03C09
33	1077	74.9	735	6	CB998867	AGENCOURT
34	1077	74.9	798	4	BG503497	602550940
35	1073.5	74.7	838	7	CK790115	AGENCOURT
36	1067	74.3	666	4	BM760772	K-EST0041
37	1067	74.3	686	5	BU664464	C1117B05
38	1067	74.3	747	4	BG527301	602565574
39	1064	74.0	874	1	AJ817172	AJ817172
40	1061	73.8	663	4	BG530210	602559316
41	1057	73.6	682	7	CN348597	170005322
42	1053	73.3	764	7	CO040283	UI-M-EMO-
43	1053	73.3	1125	2	BE788716	601475803
44	1050	73.1	872	1	AL531893	AL531893
45	1047.5	72.9	705	4	BG531406	6025559658

ALIGNMENTS

RESULT 1

LOCUS AK009584

DEFINITION Mus musculus adult male tongue cDNA, RIKEN full-length enriched library, clone:2310032N20 product:unknown EST, full insert

943 bp mRNA linear HTC 03-APR-2004

ACCESSION AK009584

VERSION AK009584.1 GI:12844467

KEYWORDS HTC; CAP-trapper.

SOURCE Mus musculus (house mouse)

ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1

Carninci, P. and Hayashizaki, Y.

High-efficiency full-length cDNA cloning

Meth. Enzymol. 303, 19-44 (1999)

PUBLISHED 10349636

REFERENCE 2

Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.

Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes

Genome Res. 10 (10), 1617-1630 (2000)

20499374

11042159

REFERENCE 3

Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P., Konno, H., Akiyama, J., Nishi, K., Kitsuai, T., Tashiro, H., Itoh, M., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujiwaka, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watahiki, M.,

GenCore version 5.1.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: April 23, 2005, 13:59:05 ; Search time 534 Seconds
(without alignments)
3137.847 Million cell updates/sec

Title: US-10-736-892-12

Perfect score: 1437

Sequence: 1 MAAGFKTVEPLEYRRFLKE.....TRHKEYSKLDEVIQSMKKH 276

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 5633728 seqs, 3035525691 residues

Total number of hits satisfying chosen parameters: 11267456

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=klp
-Q=/cgn2_1/USFTO_spool_p/US10736892/runat_22042005_115018_25420/app_query.fasta_1.455
-DB=Published Applications NA -OFMT=fastap -SUFFIX=rnpb -MINWATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blotsum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100
-THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTPMT=pto -NORM=ext -HAPSIZE=500 -MINLEN=0
-MAXLEN=2000000000 -USER=US10736892 @CGN 1 1 723 @runat 22042005 115018 25420
-NCFU=6 -ICFU=3 -NO_WMAP -LARGQUERY -NEG_SCORES=0 -WAIT -DSBLOCK=100
-LONGLOG -DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

Published Applications NA:
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq.*
2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq.*
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq.*
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq.*
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq.*
6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq.*
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq.*
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq.*
9: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq.*
10: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq.*
11: /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq.*
12: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq.*
13: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq.*
14: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq.*
15: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq.*
16: /cgn2_6/ptodata/1/pubpna/US10D_PUBCOMB.seq.*
17: /cgn2_6/ptodata/1/pubpna/US10E_PUBCOMB.seq.*
18: /cgn2_6/ptodata/1/pubpna/US10F_PUBCOMB.seq.*
19: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq.*
20: /cgn2_6/ptodata/1/pubpna/US11_NEW_PUB.seq.*
21: /cgn2_6/ptodata/1/pubpna/US16_NEW_PUB.seq.*
22: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	1201	83.6	1372	14	US-10-198-846-11485
2	937.5	65.2	2835	17	US-10-108-846-11485
3	818	56.9	520	17	US-10-242-535A-20164
4	818	56.9	520	17	US-10-085-783A-20164
C 5	714.5	49.7	1795	18	US-10-357-930-25092
6	708	49.3	532	17	US-10-242-535A-49800
7	708	49.3	532	17	US-10-085-783A-49800
C 8	544	37.9	362	11	US-09-969-034-2669
9	525	36.5	352	17	US-10-242-535A-50191
10	525	36.5	352	17	US-10-085-783A-50191
11	450	31.3	451	11	US-09-969-034-2798
12	434	30.2	435	17	US-10-242-535A-51089
13	434	30.2	435	17	US-10-085-783A-51089
C 14	429	29.9	438	10	US-09-873-367C-935
C 15	429	29.9	438	19	US-10-843-641A-935
C 16	427	29.7	1505	17	US-10-424-599-134388
C 17	410	28.5	1284	17	US-10-424-599-134389
C 18	407	28.3	384	9	US-09-917-800A-302
19	390.5	27.2	437	14	US-10-060-036-2224
20	389	27.1	1308	17	US-10-425-114-3139
21	389	27.1	1379	17	US-10-425-114-32338
22	376	26.2	1371	18	US-10-425-115-43059
23	374	26.0	1106	18	US-10-437-963-17351
24	371	25.8	1170	17	US-10-425-114-2929
25	331.5	23.1	501	9	US-09-783-590-7844
C 26	277.5	19.3	70123	13	US-10-087-192-610
27	267	18.6	1776	17	US-10-425-114-25528
28	267	18.6	1842	18	US-10-425-115-178931
29	264	18.4	1208	18	US-10-739-930-417
30	257	17.9	1148	16	US-10-119-428-54
31	255.5	17.8	675	18	US-10-767-701-7418
C 32	254.5	17.0	3740	9	US-09-925-300-345
33	244	17.0	560	11	US-09-969-034-2950
34	244	17.0	1627	18	US-10-437-963-41860
35	237.5	16.5	656	18	US-10-425-115-109522
36	237.5	16.5	869	14	US-10-198-846-2845
37	233.5	16.2	1311	17	US-10-310-154-257
C 38	232.5	16.2	1792	18	US-10-739-930-5446
39	231	16.1	535	11	US-09-969-034-4175
C 40	226	15.7	651	11	US-09-969-034-3699
41	225.5	15.7	1542	9	US-09-962-832-4
42	225.5	15.7	1542	19	US-10-843-641A-5890
43	206.5	14.4	1110	18	US-10-425-115-130039
44	202	14.1	448	18	US-10-767-701-30731
45	200	13.9	1981	17	US-10-424-599-111188

ALIGNMENTS

RESULT 1
US-10-198-846-11485/c
; Sequence 11485, Application US/10198846
; Publication No. US2003009974A1
; GENERAL INFORMATION:
; APPLICANT: Lillie, James
; APPLICANT: Xu, Yongyao
; APPLICANT: Wang, Youzhen
; APPLICANT: Steinmann, Kathleen
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS
; TITLE OF INVENTION: FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; TITLE OF INVENTION: THERAPY OF BREAST CANCER
; FILE REFERENCE: MRI-049
; CURRENT APPLICATION NUMBER: US/10198.846
; CURRENT FILING DATE: 2002-07-18
; PRIOR FILING DATE: 60/306,220
; PRIOR FILING DATE: 2001-07-18
; NUMBER OF SEQ ID NOS: 14084
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11485

```
; LENGTH: 1372
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-198-846-11485

Alignment Scores:
Pred. No.:      3,738-150      Length:      1372
Score:          1201.00      Matches:      250
Percent Similarity: 66.16%      Conservative: 10
Best Local Similarity: 63.16%      Mismatches: 16
Query Match:      83.58%      Indels:      1
DB:              14          Gaps:      1

US-10-736-892-12 (1-276) x US-10-198-846-11485 (1-1372)

QY 1 MetAlaAlaGlyPheLysThrValGluProLeuGluTyrArgArgPheLeuLysGlu 20
DB 1278 ATGGCGGCTGGGTTCAAAACCGTGAACCTCTGGAGTATTACAGAGAGATTCTGAAGAG 1219
QY 21 AsnCysArgProAspGlyArgGluLeuGlyGluPheArgThrThrValAsnIleGly 40
DB 1218 AACTCCGCTCCTGATGGAGAGAACTTGGTGAATTCAGAACCCACAACACTGTCAACATCGT 1159
QY 41 SerIleSerThrAlaAspGlySerAlaLeuValLysLeuGlyAsnThrThrValIleCys 60
DB 1158 TCATATAGTACCGCAGATGGTCTCTCTTAGGAAGTTGGGAATACATACAGTAACTCTGT 1099
QY 61 GlyValLysAlaGluPheAlaAlaProProValAspAlaProAspArgGlyTyrValVal 80
DB 1098 GGAGTTAAAGCAGAAATTTGCAGCACCATCAACAGATGCCCTGTATAAAGGATACGTTGT 1039
QY 81 ProAsnValAspLeuProLeuCysSerSerArgPheArgThrGlyProProGlyGlu 100
DB 1038 CCTAATGTGGATCTACCAACCCCTGTGTTCATCGAGATTCGGGTCTGGACCTCCTGGAGAA 979
QY 101 GluAlaGlnValThrSerGlnPheIleAlaAspValIleGluAsnSerHisIleIleLys 120
DB 978 GAGGCCCAAGTGGCTAGCAATTCATTCAGATGTCATGAAATTCACAGTAATTCAG 919
QY 121 LysGluAspLeuCysIleSerProGlyLysLeuAlaTrpValLeuTyrCysAspLeuIle 140
DB 918 AAAGAGGACTTATGCAATTTCTCCAGAAAGCTTGTCTGGTCTATATCTGTGATCTCAT 859
QY 141 CysLeuAspTyrAspGlyAsnIleLeuAspAlaCysThrPheAlaLeuAlaAlaLeu 160
DB 858 TGCCTCGACTACGATGGAGAACATTTTGGATGCCTGCACATTTGCTTCTAGCGGCTTTA 799
QY 161 LysAsnValGlnLeuProGluValThrIleAsnGluThrAlaLeuAlaGluValAsn 180
DB 798 AAAAATGTACAGTTGCTTGAAGTACTATTAATGAAGAACTGCTTTAGCAGAGTTAAT 739
QY 181 LeuLysLysLysSerTyrLeuAsnValArgAlaAsnProValAlaThrSerPheAlaVal 200
DB 738 TTAAGAGAGAAAAGTTATTGAATATTAGAACTCATCCAGTTGCAACTTCCTTGTCTGTG 679
QY 201 PheAspAspThrLeuLeuValAspProThrGlyGluGluGlyHisProValHisArg 220
DB 678 TTTGATGACACTTTCCTTATAGTTGACCTCTCTGGAGAGGAGAACATCT-GGCAACAGG 620
QY 221 AsnLeuAsnArgSerAsnGlyArgGlyArgGlnAlaValLeuSerSerGlnAla----- 238
DB 619 AACCTTAACAATAGTAATGGATGAGAGGAGCAACTCTGTGCTTTCACAAACCCAGGAT 560
QY 238 ----- 238
DB 559 GTTCCCGCCACTTCAGTCCCTAAACATATGTAGATGAAACACTCAGTGTGAGCCTTAAATTA 500
QY 238 ----- 238
DB 499 GGGAGGGCCAAATAGAAATATGGGATATTTTATTTACCTAAATTTTAGTATAGTTCTTAGC 440
QY 238 ----- 238

US-10-736-892-12 (1-276) x US-10-108-260A-1509 (1-2835)

Alignment Scores:
Pred. No.:      2,638-114      Length:      2835
Score:          937.50      Matches:      188
Percent Similarity: 89.04%      Conservative: 15
Best Local Similarity: 82.46%      Mismatches: 21
Query Match:      65.24%      Indels:      5
DB:              17          Gaps:      1

US-10-736-892-12 (1-276) x US-10-108-260A-1509 (1-2835)

QY 50 LeuValLysLeuGlyAsnThrThrValIleCysGlyValLysAlaGluPheAlaAlaPro 69
DB 2062 CTTCTTTCAATAGAGAAACATTTCTTGTCTACTA-----TAGGAATTTGCAGCACCA 2112
QY 70 ProValAspAlaProAspArgGlyTyrValValProAsnValAspLeuProLeuCys 89
DB 2113 TCAACAGATGCCCTGATAAGGATACGTTGTTCTTATGTGGATCTACCACTCCCTGTGT 2172
QY 90 SerSerArgPheArgThrGlyProGlyGluGluAlaGlnValThrSerGlnPheIle 109
DB 2173 TCATCGAGATTCGGTCTGGACCTCTGGAGAGAGGCCCAAGTGGCTAGCAATTCATT 2232
QY 110 AlaAspValIleGluAsnSerHisIleIleLysLysGluAspLeuCysIleSerProGly 129
DB 2233 GCAGATGTCAATTGAAATTCACAGATAATTCAGAAAGAGGACTTATGATTTCTCCAGGA 2292
QY 130 LysLeuAlaTrpValLeuTyrCysAspLeuIleCysLeuAspTyrAspGlyAsnIleLeu 149
DB 2293 AAGCTTGTGGTCTTACTACTGTGATCTCATTTGCTCGACTACGATGGAAAACATTTTG 2352
QY 150 AspAlaCysThrPheAlaLeuLeuAlaAlaLeuLysAsnValGlnLeuProGluValThr 169
DB 2353 GATGCTGCACATTTGCTTGTAGCGGCTTTAAAAAATGTACAGTTGCTGAAAGTTACT 2412
```




GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: April 23, 2005, 12:48:35 ; Search time 183 Seconds
(without alignments)
2467.827 Million cell updates/sec

Title: US-10-736-892-12
Perfect score: 1437
Sequence: 1 MAAGFKTVPLEYYRFLKE.....TRHKEVSKLDEVQSMKHK 276

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 1202784 seqs, 818138359 residues

Total number of hits satisfying chosen parameters: 2405568

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlp
-Q=/cgn2_1/USPTO.spool_p/US10736892/runat_22042005.115017.25344/app_query.fasta_1.455
-DB=Issued Patents NA -QMT=fastac -SUFFIX=rni -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptco -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10736892.OCGN.1.1.105 @runat_22042005.115017.25344 -NCPUS=6 -ICPU=3
-NO MMAP -LARGQUERY -NEG SCORES=0 -WAIT -DSPELOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued Patents NA:*
1: /cgn2_6/ptodata/1/ina/5A.COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B.COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A.COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B.COMB.seq:*
5: /cgn2_6/ptodata/1/ina/PTCUS.COMB.seq:*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
c 1	1213.5	84.4	3887	4	US-09-976-594-660
2	257	17.9	1058	4	US-09-976-594-325
3	187.5	13.0	678	4	US-09-270-767-1289
4	187.5	13.0	678	4	US-09-270-767-16571
5	183	12.7	1050	4	US-09-248-796A-2912
6	121	9.5	1179	4	US-09-248-796A-524
7	121	8.4	1416	4	US-09-107-532A-2051
8	100	7.0	1368	4	US-09-134-000C-2669
9	100	7.0	1920	4	US-09-252-991A-15935
c 10	99.5	6.9	2633	4	US-09-799-451-531
11	99	6.9	759	4	US-09-252-991A-7205
c 12	99	6.9	807	4	US-09-252-991A-7492

13	95.5	6.6	780	4	US-09-489-039A-3433	Sequence 3433, Ap
14	94.5	6.6	774	4	US-09-543-681A-3309	Sequence 3309, Ap
15	91.5	6.4	278	4	US-09-313-294A-4199	Sequence 4199, Ap
16	91.5	6.4	2536	4	US-09-949-016-563	Sequence 563, App
17	91.5	6.4	2536	4	US-09-949-016-5206	Sequence 5206, Ap
18	91.5	6.4	2634	4	US-09-691-538A-10	Sequence 10, Appl
19	91.5	6.4	6114	4	US-09-914-272A-4	Sequence 4, Appl
20	91.5	6.4	6114	4	US-10-638-333-4	Sequence 4, Appl
c 21	91.5	6.4	4403765	3	US-09-103-840A-2	Sequence 1, Appl
c 22	91.5	6.4	4111529	3	US-09-103-840A-1	Sequence 1, Appl
c 23	89.5	6.2	390	4	US-09-513-999C-12562	Sequence 12562, A
c 24	87.5	6.1	18551	4	US-09-902-540-1187	Sequence 1187, Ap
c 25	87	6.1	1830121	4	US-09-557-884-1	Sequence 1, Appl
c 26	87	6.1	1830121	4	US-09-643-990A-1	Sequence 1, Appl
c 27	86.5	6.0	4287	4	US-09-902-540-2590	Sequence 2590, Ap
c 28	86.5	6.0	16584	4	US-09-902-540-1119	Sequence 1119, Ap
c 29	85.5	5.9	103377	4	US-09-949-016-14089	Sequence 14089, A
c 30	85	5.9	3177	4	US-09-489-039A-4006	Sequence 4006, Ap
c 31	84	5.8	603	4	US-09-252-991A-12951	Sequence 12951, A
c 32	82.5	5.7	2825	4	US-09-949-016-1895	Sequence 1895, Ap
c 33	82.5	5.7	16802	4	US-09-949-016-12622	Sequence 12622, A
c 34	82	5.7	316	4	US-09-621-976-828	Sequence 828, App
c 35	82	5.7	735	4	US-09-328-352-1502	Sequence 1502, Ap
c 36	82	5.7	1566	4	US-09-270-767-15057	Sequence 15057, A
c 37	82	5.7	3425	4	US-09-799-451-555	Sequence 555, App
c 38	82	5.7	36470	4	US-08-311-731A-123	Sequence 123, App
c 39	81.5	5.7	1185	4	US-09-252-991A-7914	Sequence 7914, Ap
c 40	81.5	5.7	1347	4	US-09-252-991A-7642	Sequence 7642, Ap
c 41	81.5	5.7	2121	3	US-09-134-001C-2071	Sequence 2071, Ap
c 42	81.5	5.7	15297	3	US-09-817-180-3	Sequence 3, Appl
c 43	81.5	5.7	15297	4	US-10-003-295-3	Sequence 3, Appl
c 44	81.5	5.7	19152	4	US-09-949-016-12110	Sequence 12110, A
c 45	81.5	5.7	19153	4	US-09-949-016-15795	Sequence 15795, A

ALIGNMENTS

RESULT 1

US-09-976-594-660/c
; Sequence 660, Application US/09976594
; Patent No. 6673549
; GENERAL INFORMATION:
; APPLICANT: Buchbinder, Jenny
; TITLE OF INVENTION: GENES EXPRESSED IN C3A LIVER CELL CULTURES TREATED WITH STEROIDS
; FILE REFERENCE: PA-0041 US
; CURRENT APPLICATION NUMBER: US/09/976,594
; PRIOR FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: 60/240,409
; PRIOR FILING DATE: 2000-10-12
; NUMBER OF SEQ ID NOS: 1143
; SOFTWARE: PERL Program
; SEQ ID NO 660
; LENGTH: 3887
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. 6673549 033627.33
; NAME/KEY: unsure
; LOCATION: 2483, 2486
; OTHER INFORMATION: a, t, c, g, or other
US-09-976-594-660

Alignment Scores:
Pred. No.: 9.55e-142 Length: 3887
Score: 1213.50 Matches: 245
Percent Similarity: 90.75% Conservative: 10
Best Local Similarity: 87.19% Mismatches: 17
Query Match: 84.45% Indels: 11
DB: Gaps: 1

US-10-736-892-12 (1-276) x US-09-976-594-660 (1-3887)

```
QY 5 PheLysThrValGluProLeuGluTyrTyrArgArgPheLeuLysGluAsnCysArgPro 24
Db |||||
3791 TTCAAAACCGTGGAACTCTGAGTATTACAGGAGATTCTGAAAGAGAACTGCCGTCT 3732
QY 25 AspGlyArgGluLeuGlyGluPheArgThrThrValAsnIleGlySerIleSerThr 44
Db |||||
3731 GATGAAGAGAACTTGGTGAATTCAGAACCACTGTCAACATCGGTTCATTAAGTACC 3672
QY 45 AlaAspGlySerAlaLeuValLysLeuGlyAsnThrThrValIleCysGlyValLysAla 64
Db |||||
3671 GCAGATGGTCTGCTTGTAGTGAATTCAGAAATACATACAGTAATCTGTGGAGTTAAGC- 3613
QY 65 -----GluPheAlaLaProProValAspAlaProAspArg 76
Db |||||
3612 AATGACTTTTCAGTAACCTGTCAGGAATTTGCAGCACCATCAACAGATGCCCTGATAAA 3553
QY 77 GlyTyrValValProAsnValAspLeuProLeuCysSerSerArgPheArgThrGly 96
Db |||||
3552 GGATACGTTGTTCTTAATGTGATCTACCAACCCCTGTGTTCATCGAGATTCGGTCTGA 3493
QY 97 ProProGlyGluGluAlaGlnValThrSerGlnPheIleAlaAspValIleGluAsnSer 116
Db |||||
3492 CCTCTCGAGAGAGGCCCACTGCTAGCCAGTTCATTCAGAGATGTCATTGAAATTCAC 3433
QY 117 HisIleLeuLysGluAspLeuCysIleSerProGlyLysLeuAlaTirpValLeuTyr 136
Db |||||
3432 CAGATAATTCAGAAAGAGACTTATGCAATTTCTCCAGGAAAGCTTGTCTGGTCTATAC 3373
QY 137 CysAspLeuIleCysLeuAspTyrAspGlyAsnIleLeuAspAlaCysThrPheAlaLeu 156
Db |||||
3372 TGTGATCTCATTTGCTCGACTACGATGGAACATTTTGGATGCTCGCACATTTGCTTTG 3313
QY 157 LeuAlaAlaLeuLysAsnValGlnLeuProGluValThrIleAsnGluGluThrAlaLeu 176
Db |||||
3312 CTAGCGGCTTTAAATAATGTACAGTTGCTGAAGTTACTATAAATGAAGAACTGCTTTA 3253
QY 177 AlaGluValAsnLeuLysLysSerTyrLeuAsnValArgAlaAsnProValAlaThr 196
Db |||||
3252 CGAAGATTAATTTAAAGAAAGAAAGTATTTGAATATTAGAACTTCACAGTTGCAACT 3193
QY 197 SerPheAlaValPheAspThrLeuLeuIleValAspProThrGlyGluGluGlyHis 216
Db |||||
3192 TCTTTGCTGTTTGTGACACTTGTCTTATGTTGACCTTACCTGGAGAGAGGAAACAT 3133
QY 217 ProValHisArgAsnLeuAsnArgSerAsnGlyArgGlyArgGlnAlaValLeuSerSer 236
Db |||||
3132 CT-GGCAACAGGAACCTTAACAATAGTAATGATGAGGAGGCAAACTCTGTTGCTTCA 3074
QY 237 GlnAlaArgTirpGluTirpAla-AlaGlyAlaLysLeuGlnAspCysMetSerArgAlaVa 256
Db |||||
3073 CAAACAGGTGAAGTGGCTAACTGGAGCTAAACTTCAGGACTGTATGAGCCGAGCAGT 3014
QY 256 IThrArgHisLysGluValSerLysLeuAspGluValIleGlnSerMetLysHisLys 276
Db |||||
3013 TACAAGACACAAAGAAGTTAAATAACTGATGATGAGTAATTAAGATTAAGAAACCCAA 2954
QY 276 s 276
Db 2953 A 2953
```

RESULT 2

```
US-09-976-594-325
; Sequence 325, Application US/09976594
; Patent No. 6673549
; GENERAL INFORMATION:
; APPLICANT: Furness, Michael
; APPLICANT: Buchbinder, Jenny
; TITLE OF INVENTION: GENES EXPRESSED IN C3A LIVER CELL CULTURES TREATED WITH STEROIDS
; FILE REFERENCE: PA-0041 US
; CURRENT APPLICATION NUMBER: US/09/976,594
; CURRENT FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: 60/240,409
```

```
; PRIOR FILING DATE: 2000-10-12
; NUMBER OF SEQ ID NOS: 1143
; SOFTWARE: PERL Program
; SEQ ID NO 325
; LENGTH: 1058
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. 6673549 2551987CB1
US-09-976-594-325
```

Alignment Scores:

Pred. No.:	3.41e-22	Length:	1058
Score:	257.00	Matches:	83
Percent Similarity:	45.20%	Conservative:	44
Best Local Similarity:	29.54%	Mismatches:	111
Query Match:	17.88%	Indels:	44
DB:	4	Gaps:	8

US-10-736-892-12 (1-276) x US-09-976-594-325 (1-1058)

QY	18	LeuLysGluAsnCysArgProAspGlyArgGluLeuGlyGluPheArgThrThrThrVal	37
Db	87	GTCCAGGAAGACCTCCGTTGATGGCGTGTGAGGACTACCGATGTGTGCGAAGTG	146
QY	38	AsnIleGlySerIleSerThrAlaAspGlySerAlaLeuValLysLeuGlyAsnThrThr	57
Db	147	GAAACTGATGGTGTCCACACACTAGTGGTCCGCCAGGGTCAAGCTGGGTCAACAGAC	206
QY	58	ValIleCysGlyValLysAlaGluPheAlaAlaProProValAspAlaProAspArgGly	77
Db	207	ATCTTGGTGGAGTGAAGCAGAAATGGGACGCCGAGCTGGAGAAACCAAAATGAAGGC	266
QY	78	TyrValValProAsnValAspLeuProLeuCysSerSerArgPheArgThrGlyPro	97
Db	267	TACTTGGATGTTTGTGACTGTTTACCCAGTGTACCCCTGATTTGAA---GGTAGA	323
QY	98	ProGlyGluGlu---AlaGlnValThrSerGlnPheIleAlaAspValIleGluAsnSer	116
Db	324	GGAGGTGATGACCTTGGCAGCAGATCGTAACACCTCTATCGGATATTAAACAATAAA	383
QY	117	HisIleLeuLysLysGluAspLeuCysIleSerProGlyLysLeuAlaTirpValLeuTyr	136
Db	384	AGCAGTGTGCACTTAAAGACCTCTGCATTAGTCTCTCGGAGACACTGCTGGTCTCTAT	443
QY	137	CysAspLeuIleCysLeuAspTyrAspGlyAsnIleLeuAspAlaCysThrPheAlaLeu	156
Db	444	GTGGATGCTGCTTCTGGAATGGTGGAAATTTGTTGATGCCATTCCATTGCTGTA	503
QY	157	LeuAlaAlaLeuLysAsnValGlnLeuProGluVal---ThrIleAsnGluGluThrAla	175
Db	504	AAGCTGCTCTCTCAATACAAAGGATACCAAGGGTTCGAGTTTTCGAGGATGAAGAGGG	563
QY	176	LeuAlaGluValAsnLeuLysLysSerTyr-----LeuAsnValArgAla	191
Db	564	TCGAAGGACATTAAGTTGTCAGATGACCCCTTATGATGTCATACGACTAAGTGTGGAGAT	623
QY	192	AsnProValAlaThrSerPheAla-----ValPheAspAspThrLeu	205
Db	624	GTCCCTGCAATGTCACCTCTGTGCAAGATTTGGCTATCGGCATGTGGTGGATGCTACTCT	683
QY	206	-----LeuIleValAspProThrGlyGlu-GluGlyHis	216
Db	684	CAGGAGGAGGCTGCTCGCTGCCAGCTTGTGTGTCGTGTCGTCGTCAGCAGCAAGAGAG-	738
QY	216	sProValHisArgAsnLeuAsnArgSerAsnGlyArgGlyArgGlnAlaValLeuSerSe	236
Db	739	-----TTGTGACGTGTCATGAGGAAG--	759
QY	236	rGlnAlaArgTirpGluTirpAlaAlaGlyAlaLysLeuGlnAspCysMetSerArgAlaVa	256
Db	760	-----TGGGAAGGGCAGCTCGGACCCAGAGAGCATCTTCGAGATG-AT	802

XX WPI; 2002-292408/33.
 DR P-PSDB; ABB97415.
 XX
 XX An isolated polynucleotide for treating diseases associated with its
 PT encoded polypeptide such as cancer and multiple sclerosis.
 XX
 XX Claim 1; SEQ ID NO 239; 509pp; English.
 PS
 XX The present invention provides the protein and coding sequences of 444
 CC novel human proteins. These were isolated from expressed sequences tags
 CC (ESTs). They can be used to stimulate cell growth, to regulate
 CC haematopoiesis e.g. to treat aplastic anemia, to help tissue regrowth
 CC e.g. in burn treatment, to regulate the immune system e.g. to treat
 CC multiple sclerosis, to regulate activin or inhibin e.g. to treat
 CC infertility, to regulate haemostasis or thrombolysis e.g. to treat stroke
 CC and cancer, to screen for drugs, to treat inflammatory conditions e.g.
 CC rheumatoid arthritis, and to treat nervous system disorders e.g.
 CC Parkinson's disease. The present sequence is a coding sequence of the
 CC invention
 XX
 SQ Sequence 1009 BP; 328 A; 199 C; 223 G; 259 T; 0 U; 0 Other;

Alignment Scores:
 Pred. No.: 2,076-135 Length: 1009
 Score: 1269.00 Matches: 250
 Percent Similarity: 93.86% Conservative: 10
 Best Local Similarity: 90.25% Mismatches: 16
 Query Match: 88.31% Indels: 2
 DB: 6 Gaps: 0

US-10-736-892-12 (1-276) x ABB59828 (1-1009)

QY 1 MetAlaAlaGlyPheLysThrValGluProLeuGluTyrArgArgPheLeuLysGlu 20
 Db 77 ATGGCGCTGGTTCACAAACCGTGAACCTCTGGAGTATTACAGAGATTCTTGAAGAG 136
 QY 21 AonCysArgProAspGlyArgGluLeuGluPheArgThrThrValAsnIleGly 40
 Db 137 AACTGCCGCTCGTGGTGAAGAGAACTTGGTGAATTCAGAACCCCACTGTCAACATCGT 196
 QY 41 SerIleSerThrAlaAspGlySerAlaLeuValLysLeuGluValThrValIleCys 60
 Db 197 TCAATTAGTACCGAGAGTGTCTCTGTTAGTGAAGTTGGGAAATACAGTAATCTGT 256
 QY 61 GlyValLysAlaGluPheAlaAlaProProValAspAlaProAspArgGlyTyrValVal 80
 Db 257 CGAGTTAAAGCAGAAATTTGCAGCACCACATCAACAGATGCCCTGTATTAAGGATACGTTGT 316
 QY 81 ProAsnValAspLeuProProLeuCysSerArgPheArgThrGlyProGlyGlu 100
 Db 317 CCTAATGTGGATCTACACCCCTGTGTATCATCGAATTCGGTCTGGACCTCTCGGAGAA 376
 QY 101 GluAlaGlnValThrSerGlnPheIleAlaAspValIleGluAsnSerHisIleIleLys 120
 Db 377 GAGGCCCAAGTGGCTAGCCAGTTCATTGCAGATGTCATTGAAATTCACAGTAATTCAG 436
 QY 121 LysGluAspLeuCysIleSerProGlyLysLeuAlaTrpValLeuTyrCysAspIleuile 140
 Db 437 AAAGAGGAGCTTATGATTTCTCCAGAAAGCTTGTCTGGGTTCTATATCTGATCTCAT 496
 QY 141 CysLeuAspTyrAspGlyAsnIleLeuAspAlaCysThrPheAlaLeuLeuAlaLeu 160
 Db 497 TGCCTCGACTACGATGGAAACATTTGGATGCTGCACATTTGCTTCTGAGCGCTTAA 556
 QY 161 LysAsnValGlnLeuProGluValThrIleAsnGluGluThrAlaLeuAlaGluValAsn 180
 Db 557 AAAAATGTACAGTTGCCTGAAGTACTATATAATGAAGAACTGCTTTAGCAGAAAGTTAA 616
 QY 181 LeuLysLysSerTyrLeuAsnValArgAlaAsnProValAlaThrSerPheAlaVal 200
 Db 617 TTAAGAAGAAAGTATTATTGAATATTAGAACTCATCCAGTTGCAGTTCCTTGTCTGTG 676

QY 201 PheAspAspThrLeuLeuLeuValAspProThrGlyGluGluGlyHisProValHisArg 220
 Db 677 TTTGATGACACTTTTGTATTAGTTGACCTACTCGAGAGGAGGAACATCT-GGCAACAGG 735
 QY 221 AsnLeuAsnArgSerAsnGlyArgGlyArgGlnAlaValLeuSerSerGlnAlaArgTrp 240
 Db 736 AACCTTAACAATAGTAATGAGTGAAGGAGGCAAACTCTGTGTCTTCAACACAGGTGG 795
 QY 241 GluTrpAla-AlaGlyAlaLysLeuGlnAspCysMetSerArgAlaValThrArgHisLys 260
 Db 796 AAGTGGGCTAACTGGAGCTAACTTCAGGCTGTATGAGCCGAGCAGATTACAAGACAA 855
 QY 260 sGluValSerLysLeuLeuAspGluValIleGlnSerMetLysHisLys 276
 Db 856 AGAAGTTAAAAAACTGATGGATGAAGTAATTAAGATGATGAACCCAAA 904
 RESULT 2
 ADP55134
 ID ADP55134 standard; cDNA; 2161 BP.
 AC
 XX ADP55134;
 XX
 DT 18-NOV-2004 (first entry)
 XX
 XX Human PRO cDNA sequence SEQ ID NO:1110.
 DE
 XX human; PRO; immune related disease; inflammatory immune response;
 KW immune response stimulation; anti-allergic; antianemic; antiarthritic;
 KW antiasthmatic; antidiabetic; anti-inflammatory; antipsoriatic;
 KW antirheumatic; antithyroid; CNS; dermatological; gastrointestinal;
 KW haemostatic; hepatotropic; immunosuppressant; immunosuppressive; muscular;
 KW nephrotropic; neuroprotective; osteopathic; respiratory; vasotropic;
 KW virucide; gene therapy; gene; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO2004039956-A2.
 XX
 PD 13-MAY-2004.
 XX
 PF 28-OCT-2003; 2003WO-US034381.
 XX
 PR 29-OCT-2002; 2002US-0422472P.
 XX
 PA (GETH) GENENTECH INC.
 XX
 PI Aggarwal S, Clark H, Gurney AL, Schoenfeld J, Williams PM;
 PI Wood WI, Wu TD;
 XX
 DR WPI; 2004-376182/35.
 DR P-PSDB; ADP55135.
 XX
 PT New PRO polynucleotides and polypeptides, useful in useful in diagnosing
 PT and treating an immune related disease, e.g. systemic lupus
 PT erythematosus, rheumatoid arthritis, diabetes mellitus or asthma and in
 PT stimulating an immune response.
 XX
 PS Claim 2; SEQ ID NO 1110; 3009pp; English.
 XX
 CC The present invention describes an isolated PRO nucleic acid (1). Also
 CC described: (1) a vector comprising (1); (2) a host cell comprising the
 CC vector of (1); (3) a process for producing a PRO polypeptides; (4) an
 CC isolated PRO polypeptide; (5) a chimeric molecule comprising the
 CC polypeptide of (4) fused to a heterologous amino acid sequence; (6) an
 CC antibody which specifically binds to a polypeptide of (4); (7) a
 CC composition of matter comprising a polypeptide of (4); an agonist or
 CC antagonist of the polypeptide or an antibody that binds to the
 CC polypeptide in combination with a carrier; (8) an article of manufacture
 CC comprising a container, a label on the container and a composition of
 CC matter of (7); (9) a method of treating an immune related disease in a
 CC mammal; (10) a method for determining the presence of a PRO polypeptide
 CC in a sample suspected of having the polypeptide; (11) a method of
 CC diagnosing an immune related disease or an inflammatory immune response

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: April 23, 2005, 12:41:15 ; Search time 3552 Seconds
(without alignments)
3765.101 Million cell updates/sec

Title: US-10-736-892-12

Perfect score: 1437

Sequence: 1 MAAGFKTVEPLYRYRFLKE.....TRHKEVSKLDELVIQSMKHK 276

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 4708233 seqs, 24227607955 residues

Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+P2n.model -DEV=xlp
-Q/cgn2_1/USPTO_spool_p/US10736892/runat_22042005_115016_25319/app_query.fasta_1.455
-DB=genEmbl -QMT=fastcap -SUFFIX=rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTPMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10736892 -CGEN 1 1 5600 @runat_22042005_115016_25319 -NCPU=6 -ICPU=3
-NO_MMAP -LARGQUERY -NEG_SCORES=0 -WAIT -DSBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

GenEmbl.*

1: gb.ba.*

2: gb.hcg.*

3: gb.in.*

4: gb.on.*

5: gb.ov.*

6: gb.pat.*

7: gb.ph.*

8: gb.pl.*

9: gb.pr.*

10: gb.ro.*

11: gb.ats.*

12: gb.sw.*

13: gb.un.*

14: gb.vi.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1345	93.6	912	10	MMU505005 Mus muscu
2	1340	93.2	1316	10	BC059089 Mus muscu
3	1315	91.5	926	10	BC067250 Mus muscu
4	1269	88.3	986	9	BC020773 Homo sapi

5	1269	88.3	1009	6	AX405824	Sequence
6	1247	86.8	944	9	AF025438	Homo sapi
c 7	1213.5	84.4	3887	6	AR448005	Sequence
8	1195.5	83.2	933	6	AX053242	Sequence
9	1174.5	81.7	164637	2	AC142180	Rattus no
10	1034	72.0	1132	5	EX935619	Gallus ga
11	988	68.8	993	5	CR760459	Xenopus t
12	988	68.8	1007	5	BC080360	Xenopus t
13	981	68.3	798	9	AF279901	Homo sapi
14	979	68.1	1045	5	BC041271	Xenopus 1
15	937.5	65.2	2835	6	AX834385	Sequence
16	937.5	65.2	2835	6	AX834385	Sequence
17	937	65.2	987	5	AF046875	Homo sapi
18	906	63.0	2428	9	HSMB00526	Homo sapi
19	877	61.0	636	6	AX053311	Sequence
20	877	61.0	636	6	AX053404	Sequence
21	818	56.9	520	6	CQ675238	Sequence
22	766.5	53.3	816	6	CQ732109	Sequence
c 23	714.5	49.7	1795	6	CQ493225	Sequence
c 24	713	49.6	172033	9	AL136131	Human DNA
25	713	49.6	193332	2	AL355349	Sequence
26	708	49.3	532	6	CQ704874	Homo sapi
c 27	644.5	44.9	687	6	AX053312	Sequence
c 28	644.5	44.9	687	6	AX053405	Sequence
29	621	43.2	2189	6	CQ849579	Sequence
30	621	43.2	2189	9	AK126602	Homo sapi
31	525	36.5	352	6	CQ705265	Sequence
32	508	35.4	592	5	EX930480	Gallus ga
33	434	30.2	435	6	CQ706163	Sequence
34	431	30.0	909	8	BT004614	Arabidops
c 35	429	29.9	438	6	AX330426	Sequence
c 36	420	29.2	20000	8	SPBC17D1	S.pombe c
c 37	407	28.3	384	6	AX400626	Sequence
38	394	27.4	1130	8	AK102900	Oryza sat
39	394	27.4	1254	8	AK073252	Oryza sat
40	392	27.3	1271	8	AK105427	Oryza sat
41	386	26.9	163831	2	AC150948	Bos tauru
c 42	379.5	26.4	14506	1	AE010257	Pyrococcu
c 43	360.5	25.1	319000	1	AP000006	Pyrococcu
44	359	25.0	110000	6	BD430793	Continuation (6 of
45	359	25.0	293250	1	CNSPAX02	Continuation (6 of

ALIGNMENTS

RESULT 1
MMU505005
LOCUS
DEFINITION Mus musculus mRNA for CBP-interacting protein 3 (CIP3 gene).
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
FEATURES
source

MMU505005 912 bp mRNA linear ROD 09-AUG-2002
AJ505005
AJ505005.1 GI:22208742
CBP-interacting protein 3; CIP3 gene.
Mus musculus (house mouse)
Mus musculus
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
Heery,D.M., Harries,J.C., Kindle,K.B., Viskaduraki,M., Matsuda,S., Sheppard,H.M. and Davis,A.
Diverse nuclear proteins compete with SRC1 for interaction with CREB binding protein
Unpublished
2 (bases 1 to 912)
Heery,D.M.
Direct Submission
Submitted (05-AUG-2002) Heery D.M., Biochemistry, University of Leicester, University Road, Leicester LE1 7RH, UNITED KINGDOM
Location/Qualifiers
1. .912
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/tissue_type="embryo"

```

gene      1. 912
/ gene="CIP3"
CDS       1. 831
/ gene="CIP3"
/ function="putative transcriptional regulator"
/ codon_start=1
/ product="CBP-interacting protein 3"
/ protein_id="CAD43467.1"
/db_xref="GI:22208743"
ADGSAALVIGNTTTCGKAEPAAPVDPAPGVVNVNDLPPLCSSFRFGPPGEEA
QVTSQFADVDSNVISKEDICISPGKAWLYCDLCLDVGNDILDCATFALLAAL
KNVQLPEVINEETALEAENLKKSYLVNRNPVATSEAVPDDTLIIYDPTGEEHLS
TGTLLVTDDEGKLCCLHKKPGSGLTGAKIQDCMSRAVTRHKEVSKLDEVIQSMRHK
"
832. 912
/ gene="CIP3"

3' UTR
ORIGIN
Alignment Scores:
Pred. No.: 6,8e-123 Length: 912
Score: 1345.00 Matches: 264
Percent Similarity: 97.11% Conservative: 5
Best Local Similarity: 95.31% Mismatches: 7
Query Match: 93.60% Indels: 2
DB: 10 Gaps: 0

US-10-736-892-12 (1-276) x MMU505005 (1-912)

Qy 1 MetAlaAlaGlyPheLysThrValGluProLeuGluTyrTyrArgPheLeuLysGlu 20
Db 1 ATGGCGGCTGGGTTTCAAACTGTGAGCGCTGGAGTATTACAGAGATTCTTGAAGAA 60

Qy 21 AenCysArgProAspGlyArgGluLeuGlyGluPheArgThrThrValAsnIlegly 40
Db 1 AACTGCGGCTCGATGGAAGAACTTGGTGAATTTCAGAGCCACCACTGTCAACATAGT 120

Qy 41 SerIleSerThrAlaAspGlySerAlaLeuValLysLeuLysValLysThrValIleCys 60
Db 121 TCAATCAGTACGACCGATGGCTCTGCTCTCGTGAAGCTGGGAAATACACCGTCAATTTGT 180

Qy 61 GlyValLysAlaGluPheAlaAlaProProValAspAlaProAspArgGlyTyrValVal 80
Db 181 GCGCTTAAGCAGATTTCAGACACCACTAGATGCCCCGATAGAGATATGCTGTC 240

Qy 81 ProAsnValAspLeuProLeuCysSerSerArgPheArgThrGlyProProGlyGlu 100
Db 241 CCTAATGTGATCTACACCGCTGTGTTCTTCGAGGTTCCGACAGGACCTCTCGGAA 300

Qy 101 GluAlaGlnValThrSerGlnPheIleAlaAspValIleGluAsnSerHisIleIleLys 120
Db 301 GAGGCTCAAGTAACAGCAGTTCATTGCGATGTCGATGTCGTGACAACTCACAGTGATTAG 360

Qy 121 LysGluAspLeuCysIleSerProGlyLysLeuAlaTppValLeuTyrCysAspLeuIle 140
Db 361 AAAGAGGATTATGATTTCTCCAGCAGCTTGGTTGGTTCTATATCTGTGACCTCAT 420

Qy 141 CysLeuAspTyrAspGlyAsnIleLeuAspAlaCysThrPheAlaLeuAlaLeu 160
Db 421 TGTCTAGACTACGATGGGAACATTTCGATGCTGCACATTTCCTTTGTTAGCAGCTTA 480

Qy 161 LysAsnValGlnLeuProGluValThrIleAsnGluThrAlaLeuAlaGluValAsn 180
Db 481 AAGAATGTACAGTTCCTGTAAGTTACTATAACGAAGAGACTGCTTTTAGCCGAATCAAT 540

Qy 181 LeuLysLysLysSerTyrLeuAsnValArgAlaAsnProValAlaThrSerPheAlaVal 200
Db 541 TTAAGAGAGAAAAGTTATTGATGTAGAACAAACCCAGTTGCTACTTCTTCTGCTGTG 600

Qy 201 PheAspAspThrLeuLeuIleValAspProThrGlyGluGluGlyHisProValHisArg 220
Db 601 TTTGTATGACACTTACTATAGTATCTTACCGCGAGGA-GGAGACACCTGTCTCCACAGG 659

```

221 AsnLeuAsnArgSerAsnGlyArgGlnAlaValLeuSerSerGlnAlaArgTrp 240
 660 AACCTTAACCTAGTACGAGCAGGACGCAAGCTGTCTTCAACAGCCAGGTGG 719
 241 GluTppAla-AlaGlyAlaLysLeuGlnAspCysMetSerArgAlaValThrArgHisLys 260
 720 GAGTGGGCTGACTGGAGCTAACTTCAGACTGATGAGTGGGGTAAACGAGACAA 779
 260 sGluValSerLysLeuLeuAspGluValIleGlnSerMetLysHisLys 276
 780 AGAAGTACGACCAAGCTACTGATGAAGTAATTCAGAGCATGAGACAA 828

BC059089 1316 bp mRNA linear ROD 30-JUN-2004
 Mus musculus exosome component 8, mRNA (cdna clone MGC:70125
 IMAGE:6509820), complete cds.
 BC059089.1 GI:37598849
 Mus musculus (house mouse)
 MGC.
 Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 1316)
 Strausberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G.,
 Klausner, R.D., Collins, F.S., Wagner, L., Shenmen, C.M., Schuler, G.D.,
 Altschul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K.,
 Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, F.,
 Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L.,
 Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L.,
 Scheerz, I.E., Brownstein, M.J., Usdin, T.B., Toshiyuki, S.,
 Carninci, P., Prange, C., Raha, S.S., Loquellano, N.A., Peters, G.J.,
 Abramson, R.D., Mullahy, S.J., Bosak, S.A., McSwan, P.J.,
 McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S.,
 Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W.,
 Villalón, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A.,
 Fahey, J., Helton, E., Kettner, M., Madan, A., Rodriguez, S.,
 Sanchez, A., Whitting, M., Madan, A., Young, A.C., Shevchenko, Y.,
 Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D.,
 Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M.,
 Butterfield, Y.S., Krzywinski, M.I., Skalska, U., Smalish, D.E.,
 Schnerch, A., Schein, J.E., Jones, S.J. and Marra, M.A.
 Generation and initial analysis of more than 15,000 full-length
 human and mouse cDNA sequences
 Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
 12477932
 2 (bases 1 to 1316)
 Strausberg, R.
 Direct Submission
 Submitted (01-OCT-2003) National Institutes of Health, Mammalian
 Gene Collection (MGC), Cancer Genomics Office, National Cancer
 Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
 USA
 NIH-MGC Project URL: <http://mgc.nci.nih.gov>
 Contact: MGC help desk
 Email: cgapsb-remail.nih.gov
 Tissue Procurement: Dr. David Rowe
 cDNA Library Preparation: Invitrogen Corp
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
 DNA Sequencing by: Genome Sequence Centre,
 BC Cancer Agency, Vancouver, BC, Canada
info@bcgsc.bc.ca
 Steve Jones, Sarah Barber, Mabel Brown-John, Yaron Butterfield,
 Andy Chan, Steve S. Chand, William Chow, Allison Cloutier, Ruth
 Featherstone, Malachi Griffith, Obi Griffith, Ran Guin, Nancy Liao,
 Kim MacDonald, Amara Masson, Mike R. Mayo, Josh Moran, Ryan Morin,
 Teika Olson, Diana Palmquist, Anca Petrescu, Anna Liisa Prabhui,
 Parvaneh Saedi, JR Santos, Angélique Scherch, Ursula Skalska,
 Duane Smalish, Jeff Stott, Miranda Tsai, George Yang, Jacques
 Schein, Asim Siddiqui, Rob Holt, Marco Marra.

Clone distribution: MGC clone distribution information can be found

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 23, 2005, 12:16:27 ; Search time 120 Seconds
(without alignments)
1177.782 Million cell updates/sec

Title: US-10-736-892-12

Perfect score: 1437

Sequence: 1 MAAGFTVPLEYRRFLKE.....TRHKEVSKLLDEVQSMKHK 276

Scoring table:
BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1612378 seqs, 512079187 residues

Total number of hits satisfying chosen parameters: 1612378

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Uniprot 03:.*
1: uniprot_sprot:.*
2: uniprot_trembl:.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	1227	85.4	276	1 RR43 MOUSE	Q9d753 mus musculus
2	1222	85.0	276	2 Q6PCW8	Q6pcw8 mus musculus
3	1197	83.3	270	2 Q6NX62	Q6nx62 mus musculus
4	1184	82.4	276	1 RR43 HUMAN	Q96b26 homo sapien
5	985	68.5	276	2 Q66KK0	Q66kk0 xenopus tro
6	978	68.1	276	2 Q8AVT6	Q8avt6 xenopus lae
7	918	63.9	277	2 Q6DRN4	Q6drn4 brachydanio
8	580	40.4	280	2 Q7QDD9	Q7qdd9 anopheles g
9	459.5	32.0	270	1 RR43 SCHPO	Q10205 schizosacch
10	431	30.0	302	2 Q9ZU14	Q9zul4 arabidopsis
11	379.5	26.4	277	1 ECX2 PYRFU	Q8u0m0 pyrococcus
12	360.5	25.1	274	1 ECX2 PYRHO	C59224 Pyrococcus
13	359	25.0	274	1 ECX2 PYRAB	P59118 Pyrococcus
14	333.5	23.2	271	1 ECX2 METHT	Q26778 methanobact
15	331	23.0	275	1 ECX2 SULTO	Q975g9 sulfolobus
16	325	22.6	275	1 ECX2 SULSO	Q9uxc0 sulfolobus
17	323	22.5	260	2 Q6LJ24	Q6lj24 thermophilus
18	322.5	22.4	260	1 ECX2 THEAC	Q9hip1 thermoplasm
19	321	22.3	257	2 Q647V7	Q647v7 uncultured
20	318	22.1	257	2 Q64B51	Q64b51 uncultured
21	316.5	22.0	267	1 ECX2 METKA	Q8tyc2 methanopyru
22	316	22.0	257	2 Q64D54	Q64d54 uncultured
23	313	21.8	259	1 ECX2 ARCFU	Q29756 archaeoglob
24	310.5	21.6	260	1 ECX2 THERVO	Q97b24 thermoplasm
25	310.5	21.6	276	1 ECX2 AERPE	Q9yc05 aeropyrum p
26	308.5	21.5	274	1 ECX2 PYRAE	Q8vzn0 pyrobaculum
27	300	20.9	260	2 Q64CA6	Q64ca6 uncultured
28	297.5	20.7	260	2 Q64G01	Q64g01 uncultured
29	295.5	20.6	260	2 Q6M2C0	Q6m2c0 uncultured
30	294	20.5	266	1 ECX2 METMA	Q9ptc7 methanosarc
31	288.5	20.1	395	2 Q69T30	Q69t30 oryza sativ

32	288	20.0	266	1 ECX2 METAC	Q8tqx5 methanosarc
33	279	19.4	238	2 Q64DM5	Q64dm5 uncultured
34	271.5	18.9	305	1 RR45 YEAST	Q05636 saccharomyc
35	269.5	18.8	431	2 Q8GQJ3	Q8gqj3 xenopus lae
36	264	18.4	307	2 Q8LGD7	Q8lgd7 arabidopsis
37	264	18.4	307	2 Q9LDM2	Q9ldm2 arabidopsis
38	260.5	18.1	305	2 Q75F52	Q75f52 ashbya goss
39	260.5	18.1	438	1 RR45 MOUSE	Q9jhi7 m exosome c
40	260	18.1	438	2 Q9M209	Q9m209 arabidopsis
41	257	17.9	291	1 RR42 HUMAN	Q15024 homo sapien
42	256.5	17.8	305	2 Q6CTP4	Q6ctp4 kluyveromyc
43	255	17.7	286	2 Q6ZQJ0	Q6zqj0 mus musculu
44	254.5	17.7	439	2 Q86Y48	Q86y48 homo sapien
45	254.5	17.7	456	2 Q86Y41	Q86y41 homo sapien

ALIGNMENTS

RESULT 1
RR43_MOUSE STANDARD; PRT; 276 AA.
AC Q9D753;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 05-JUL-2004 (Rel. 44, Last annotation update)
DE Exosome complex exonuclease RRP43 (EC 3.1.13.-) (Ribosomal RNA
processing protein 43) (Exosome component 8).
GN Name=Exosc8; Synonym=RRP43;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Testis; DOI=10.1038/nature01266;
RX MEDLINE=22354683; PubMed=12466851;
RA Okazaki Y., Furuno M., Kasukawa T., Adachi J., Bono H., Kondo S.,
RA Nikaido I., Osato N., Saito R., Suzuki H., Yamana I., Kiyosawa H.,
RA Yagi K., Tomaru Y., Hasegawa Y., Nogami A., Schonbach C., Gojohori T.,
RA Baldarelli R., Hill D.P., Bult C., Hume D.A., Quackenbush J.,
RA Schriml L.M., Kanapin A., Matsuda H., Batalov S., Beisel K.W.,
RA Blake J.A., Bradt D., Bruscia V., Chothia C., Corbani L.E., Cousins S.,
RA Dalla E., Dragani T.A., Fletcher C.F., Forrester A., Frazer K.S.,
RA Gaasterland T., Gariboldi M., Gissi C., Godzik A., Gough J.,
RA Grimmond S., Gustincich S., Hirokawa N., Jackson I.J., Jarvis E.D.,
RA Kanai A., Kawaji H., Kawasawa Y., Kedzierski R.M., King B.L.,
RA Konagaya A., Kurochkin I.V., Lee Y., Lenhard B., Lyons P.A.,
RA Maglott D.R., Maltais L., Marchionni L., McKenzie L., Miki H.,
RA Nagashima T., Numata K., Okido T., Pavan W.J., Pertea G., Pesole G.,
RA Petrovsky N., Pillai R., Pontius J.U., Qi D., Ramachandran S.,
RA Ravasi T., Reed J.C., Reid J., Ring B.Z., Ringwald M.,
RA Sadelin A., Schneider C., Semple C.A., Setou M., Shimada K.,
RA Sultana R., Takenaka Y., Taylor M.S., Teasdale R.D., Tomita M.,
RA Verardo R., Wagner L., Wahlestedt C., Wang Y., Watanabe Y., Wells C.,
RA Wilming L.G., Wymshaw-Boris A., Yanagisawa M., Yang I., Yang L.,
RA Yuan Z., Zavalon M., Zhu Y., Zimmer A., Carninci P., Hayatsu N.,
RA Hirozane-Kishikawa T., Konno H., Nakamura M., Sakazume N., Sato K.,
RA Shiraki T., Waki K., Kawai J., Aizawa K., Arakawa T., Fukuda S.,
RA Hara A., Hashizume W., Inotani K., Ishii Y., Itoh M., Kagawa I.,
RA Miyazaki A., Sakai K., Sasaki D., Shibata K., Shinagawa A.,
RA Yasunishi A., Yoshino M.,
RA Birney E., Hayashizaki Y.;
RT "Analysis of the mouse transcriptome based on functional annotation of
60,770 full-length cDNAs";
RL Nature 420:563-573(2002).
CC -!- FUNCTION: Component of the nuclear exosome 3'->5' exoribonuclease
complex. Required for the 3' processing of the 7S pre-RNA to the
mature 5.8S rRNA. Has a 3'-5' exonuclease activity (By
similarity).
CC -!- SUBUNIT: Component of the exosome multi-enzyme ribonuclease complex
composed of at least 11 proteins: RRP4, RRP40, RRP41/SKI6, RRP42,
RRP43, RRP44/DIS3, PM/SCI-75, RRP46, CSL4 and PM/SCI-100 (only in

the nuclear complex). Also associated with the GPase Ran (By similarity).

-1- SUBCELLULAR LOCATION: Cytoplasmic and nuclear; nucleolar (By similarity).

-1- SIMILARITY: Belongs to the RNase PH family.

THIS SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; AK009584; -; NOT ANNOTATED_CDS.

MGB; MGI:1916889; 2310032N20Rik.

InterPro; IPR001247; 3 ExonNase.

Pfam; PF01138; RNase PH; 1.

Pfam; PF03725; RNase PH C; 1.

Exonuclease; Exosome; Hydrolase; Nuclear protein; Nuclease;

RNA-binding; rRNA processing.

SEQUENCE 276 AA; 29949 MW; 297E1E45P5C0F794 CRC64;

Query Match 85.4%; Score 1227; DB 1; Length 276;
Best Local Similarity 87.3%; Pred. No. 3.5e-99;
Matches 241; Conservative 6; Mismatches 29; Indels 0; Gaps 0;

Qy 1 MAAGFKTVELEYRFLKENCRCPDGRELGEFRATTNNIGSISTADGSALVKGNTTVC 60
Db 1 MAAGFKTVELEYRFLKENCRCPDGRELGEFRATTNNIGSISTADGSALVKGNTTVC 60

Qy 61 GVKAFAAPVDAPDRGVVNPVNDLPPLCSSRFRTPGPGEEAQTSTFIADVIENSHIK 120
Db 61 GVKAFAAPVDAPDRGVVNPVNDLPPLCSSRFRTPGPGEEAQTSTFIADVDNSQVIK 120

Qy 121 KEDLCISPGKLAWLVCYDLICLDYDGNILDACTFALLAALKNVQLPEVTINEETALAEVN 180
Db 121 KEDLCISPGKLAWLVCYDLICLDYDGNILDACTFALLAALKNVQLPEVTINEETALAEVN 180

Qy 181 LKKSXLNVRNPVATSPAVFDDTLIVDPTEGEGHPVHRNLNRSNGRQAVLSSQARW 240
Db 181 LKKSXLNVRNPVATSPAVFDDTLIVDPTEGEGHLSGTGLTVVTTDDEGKLCCLHPGG 240

Qy 241 EWAAGAKLQDCMSRAVTRHKEVSKLLDEVIOQSMKHK 276
Db 241 SGLTGAKLQDCMSRAVTRHKEVSKLLDEVIOQSMRHK 276

RESULT 2

Q6PCW8 PRELIMINARY; PRT; 276 AA.

AC Q6PCW8; PRELIMINARY; PRT; 276 AA.

DT 05-JUL-2004 (TrEMBLrel. 27, Created)

DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)

DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)

DE CBP-interacting protein 3.

GN Name=Exosc8;

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OX NCBI_TaxID=10090;

RP [1]

RP SEQUENCE FROM N.A.

RC TISSUE=Limb;

RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.2426038999;

RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G., Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D., Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K., Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F., Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L., Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.B., Brownstein M.J., Udén T.B., Toshiyuki S., Carninci P., Prange C., Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,

RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H., Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W., Villallon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A., Fahy J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A., Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G., Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C., Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S., Krzywinski M.I., Skalska U., Smailus D.E., Schnerch A., Schein J.E., Jones S.J., Marra M.A.;

"Generation and initial analysis of more than 15,000 full-length human RT and mouse cDNA sequences."

RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).

RN [2]

RP SEQUENCE FROM N.A.

RC TISSUE=Limb;

RA Strausberg R.;

RL Submitted (OCT-2003) to the EMBL/GenBank/DBJ databases.

DR EMBL; BC059089; AAH59089.1; -

DR GO; GO:0000175; F:3'-5'-exoribonuclease activity; IEA.

DR GO; GO:0003723; F:RNA binding; IEA.

DR GO; GO:0006396; P:RNA processing; IEA.

DR InterPro; IPR001247; 3 ExonNase.

DR Pfam; PF01138; RNase PH; 1.

DR Pfam; PF03725; RNase PH C; 1.

SQ SEQUENCE 276 AA; 29919 MW; 3264AFEEEF714794 CRC64;

Query Match 85.0%; Score 1222; DB 2; Length 276;
Best Local Similarity 87.0%; Pred. No. 9.7e-99;
Matches 240; Conservative 6; Mismatches 30; Indels 0; Gaps 0;

Qy 1 MAAGFKTVELEYRFLKENCRCPDGRELGEFRATTNNIGSISTADGSALVKGNTTVC 60
Db 1 MAAGFKTVELEYRFLKENCRCPDGRELGEFRATTNNIGSISTADGSALVKGNTTVC 60

Qy 61 GVKAFAAPVDAPDRGVVNPVNDLPPLCSSRFRTPGPGEEAQTSTFIADVIENSHIK 120
Db 61 GVKAFAAPVDAPDRGVVNPVNDLPPLCSSRFRTPGPGEEAQTSTFIADVDNSQVIK 120

Qy 121 KEDLCISPGKLAWLVCYDLICLDYDGNILDACTFALLAALKNVQLPEVTINEETALAEVN 180
Db 121 KEDLCISPGKLAWLVCYDLICLDYDGNILDACTFALLAALKNVQLPEVTINEETALAEVN 180

Qy 181 LKKSXLNVRNPVATSPAVFDDTLIVDPTEGEGHPVHRNLNRSNGRQAVLSSQARW 240
Db 181 LKKSXLNVRNPVATSPAVFDDTLIVDPTEGEGHLSGTGLTVVTTDDEGKLCCLHPGG 240

Qy 241 EWAAGAKLQDCMSRAVTRHKEVSKLLDEVIOQSMKHK 276
Db 241 SGLTGAKLQDCMSRAVTRHKEVSKLLDEVIOQSMRHK 276

RESULT 3

Q6NX62 PRELIMINARY; PRT; 270 AA.

AC Q6NX62; PRELIMINARY; PRT; 270 AA.

DT 05-JUL-2004 (TrEMBLrel. 27, Created)

DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)

DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)

DE Exosc8 protein (fragment).

GN Name=Exosc8;

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OX NCBI_TaxID=10090;

RP [1]

RP SEQUENCE FROM N.A.

RC TISSUE=Mammary gland;

RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.2426038999;

RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G., Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D., Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K., Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F., Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L., Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.B., Brownstein M.J., Udén T.B., Toshiyuki S., Carninci P., Prange C., Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 23, 2005, 12:17:22 ; Search time 44 Seconds
(without alignments)
603.542 Million cell updates/sec

Title: US-10-736-892-12
Perfect score: 1437
Sequence: 1 MAAAGFTVBLEYRFLKE.....TRHKEVSKLLDEVQSMKHK 276
Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues
Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR 79: *
1: pir1: *
2: pir2: *
3: pir3: *
4: pir4: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	459.5	32.0	270	2 S67389	conserved hypotet
2	431	30.0	302	2 D96625	hypothetical prote
3	360.5	25.1	274	2 D71032	probable autoantig
4	359	25.0	274	2 F75181	polyribonucleotide
5	333.5	23.2	271	2 H69190	conserved hypotet
6	325	22.6	275	2 G90221	conserved hypotet
7	313	21.8	259	2 F69311	conserved hypotet
8	310.5	21.6	276	2 E72623	probable autoantig
9	271.5	18.9	305	2 S70136	probable membrane
10	260	18.1	438	2 T47861	nucleolar autoanti
11	253.5	17.6	617	2 T28842	hypothetical prote
12	250.5	17.4	291	2 T41599	75K autoantigen ho
13	182.5	12.7	574	2 T16328	hypothetical prote
14	168.5	11.7	372	2 G01425	nucleolar 75K auto
15	157	10.9	305	2 T21601	hypothetical prote
16	148.5	10.3	299	2 T39602	conserved hypotet
17	143.5	10.0	394	2 S12917	hypothetical prote
18	123.5	8.6	245	1 A44914	tRNA nucleotidyltr
19	121	8.4	239	2 G97821	tRNA nucleotidyltr
20	119	8.3	240	2 C71668	ribonuclease ph (r
21	115.5	8.0	241	2 T47954	exonuclease RRP41
22	115	8.0	248	2 H90221	ribonuclease PH (r
23	111.5	7.8	246	2 D72623	probable ribonucle
24	111	7.7	257	2 D84033	ribonuclease PH rp
25	110.5	7.7	248	2 AF1229	ribonuclease PH ho
26	109.5	7.6	248	2 AH1582	ribonuclease PH ho
27	108	7.5	242	2 T11646	tRNA nucleotidyltr
28	104	7.2	242	2 T11740	hypothetical prote
29	102	7.1	545	1 JX0225	cytochrome P450 CY

30	99.5	6.9	257	2 D82350	ribonuclease PH VC
31	98.5	6.9	238	1 Q9ECPE	tRNA nucleotidyltr
32	98.5	6.9	238	2 AG0970	RNase PH [imported
33	97.5	6.8	238	2 G86040	RNase PH [imported
34	97.5	6.8	238	2 F91193	RNase PH [imported
35	95.5	6.6	245	2 T36127	probable ribonucle
36	94.5	6.6	498	2 A97061	uncharacterized pr
37	94	6.5	240	2 A69191	tRNA nucleotidyltr
38	91.5	6.4	1872	2 T00339	hypothetical prote
39	90.5	6.3	3839	2 T49799	related to TOM1 pr
40	90	6.3	238	2 AI3473	tRNA nucleotidyltr
41	90	6.3	255	2 B70380	tRNA nucleotidyltr
42	89.5	6.2	239	2 D82978	ribonuclease PH PA
43	89.5	6.2	249	2 E71032	hypothetical prote
44	89	6.2	258	2 E69311	ribonuclease PH (r
45	88.5	6.2	239	2 T49262	hypothetical prote

ALIGNMENTS

RESULT 1

S67389
conserved hypothetical protein SPBC17D1.03c - fission yeast (Schizosaccharomyces pombe)
C;Species: Schizosaccharomyces pombe
C;Date: 19-Jul-1996 #sequence_revision 13-Mar-1997 #text_change 09-Jul-2004
C;Accession: T39706; S67389
R;Wood, V.; Rajandream, M.A.; Barrell, B.G.; Beck, A.; Reinhardt, R.
submitted to the EMBL Data Library, August 1998
A;Reference number: Z21871
A;Accession: T39706
A;Status: translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-270 <WOO>
A;Cross-references: UNIPROT:Q10205; EMBL:AL031322; PIDN:CAA20427.1; GSPDB:GN00067; SPDB:1
A;Experimental source: strain 972h-; cosmid cl7D1
C;Genetics:
A;Gene: SPBC17D1.03c
A;Map position: 2
A;Introns: 15/2; 127/2
C;Superfamily: conserved hypothetical protein MTH682

Query Match	32.0%;	Score	459.5;	DB	2;	Length	270;
Best Local Similarity	42.9%;	Pred. NO.	9.7e-34;				
Matches	93;	Conservative	41;	Mismatches	68;	Indels	15;
						Gaps	3;
QY	5	FKTVELEYRRLKENC	PDGRELGEFRITTVNIGSISTADGSAVLKLTNTTVCVKA	64			
DB	17	FKKITPEQYLSHLNQDVS	SGRSVSEFRREIVINDNCISTANGSAIRAGENVFVCGIKA	76			
QY	65	EFAAPPVDA	PDGYYVVPNDLPPLCSSRFRTPGPGEEAOVTQFIADVIENSHIHKEDL	124			
DB	77	ETAEPPENFNEG	IVPNLESLPLCSKPKPGPSDLAQVVSSELHQTLOQSNLINQSL	136			
QY	125	CISPGKGLAWLYCDL	ICLDYDGNILDACFPALLAALKNVQLPEVTINEETALAEVNLKKK	184			
DB	137	CIFEKKAAMWLYADI	ICLVYDGSADFVAAALFAALKTKLP-----TAVWDEDLERV	189			
QY	185	SYLVNRPV	-----ATSFVFDPTLLIVDTGEE	214			
DB	190	ICASTLTRPVQLST	EVRSFWSVFDKLL-ADPTDEE	225			

RESULT 2

D96625
hypothetical protein T2K10.14 [imported] - Arabidopsis thaliana
C;Species: Arabidopsis thaliana [mouse-ear cress]
C;Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 09-Jul-2004
C;Accession: D96625
R;Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso, J.; Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Dewar, K.; Creasy, T.H.; Huizar, L.; ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000

A;Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziali,
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A;Authors: Salberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A;Title: Sequence and analysis of chromosome 1 of the plant *Arabidopsis*.
A;Reference number: A86141; MUID:21016719; PMID:11130712
A;Accession: D96625
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-302 <STO>
A;Cross-references: UNIPROT:Q9ZUI4; GB:AE005173; NID:g4249388; PIDN:AAD14485.1; GSPDB:GN
C;Genetics:
A;Map position: 1
C;Superfamily: conserved hypothetical protein MTH682

Query Match 30.0%; Score 431; DB 2; Length 302;
Best Local Similarity 32.8%; Pred. No. 4.2e-31;
Matches 95; Conservative 55; Mismatches 106; Indels 34; Gaps 4;
Qy 5 FKTEPELYRRFLKENCRRPDGRELGEFRRTTNIGSISTADGSAIVKLGNTTIVICGVA 64
Db 19 FRRIFFLRFERHLSERLPDQGLKARDIVNLGLVSTADGSAIKGISTTTLAAIRM 78
Qy 65 EFAAPPVDAPRGVYVNVDPPLCSSRFRTPGPEEAQVTSQFIADVIENSHIIKKEDL 124
Db 79 EWTFSTDSPEGCIATEFHPPICSPVTRGRFAEAPVSKLSLSTILSSGMDIKEL 138
Qy 125 CISPGKLAWLVCILCLDYDGNILDACTFALLAALKNVQLPEVTINEETALAEVN---- 180
Db 139 CLVSGKAAMGYLDIYCLDADGALFDALAAVAFAFSLQIPIVALNDNGRIVAVTGEKD 198
Qy 181 -----LKKSVILNVRANPVATSPAVFDLTLIVDPTGEGHPVHRNLNRSNGR 228
Db 199 QDNALITEKEAVNKEKRLTKNIPFSLT-CILHKNVILADPTTEESIMDTLVT----- 252
Qy 229 GRQAVLSQARW-----EWAAGAKLODCMSRAVTRHKEVSKLLDEV 269
Db 253 ---VVLSDSDQWVFKSGGAALAYSSAISKVELARKAKELKQILGEM 299

RESULT 3
D71032
Probable autoantigen like protein - Pyrococcus horikoshii
C;Species: Pyrococcus horikoshii
C;Date: 14-Aug-1998 #sequence_revision 14-Aug-1998 #text_change 12-Jul-2004
C;Accession: D71032
R;Kawarabayashi, Y.; Sawada, M.; Horikawa, H.; Haikawa, Y.; Hino, Y.; Yamamoto, S.; Sekin
M.; Ohfuku, Y.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kushida, N.; Oguchi
DNA Res. 5, 55-76, 1998
A;Title: Complete sequence and gene organization of the genome of a hyper-thermophilic a
A;Reference number: A71000; MUID:98344137; PMID:9679194
A;Accession: D71032
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-274 <KAW>
A;Cross-references: UNIPROT:O59224; GB:AP0000006; NID:g3236133; PIDN:BAA30660.1; PID:g325
A;Experimental source: strain O73
A;Note: this accession replaces an interim accession for a sequence replaced by GenBank
C;Genetics:
A;Gene: PH1548

Query Match 25.1%; Score 360.5; DB 2; Length 274;
Best Local Similarity 33.5%; Pred. No. 8.3e-25;
Matches 91; Conservative 46; Mismatches 100; Indels 35; Gaps 5;
Qy 18 LKENCRRPDGRELGEFRRTTNIGSISTADGSAIVKLGNTTIVICGKAEFAAPPVADAPRG 77
Db 19 LKEGKRIDRGFDYRPIEIVGLIEKAGSALVKGSTQVLVGIKITLGEFFPTPNMG 78
Qy 78 YVNVNVDLPPLCSSRFRTPGPEEAQVTSQFIADVIENSHIIKKEDLCISPGKLAWLVC 137

Db 79 VMTTNVELVLASPTTFEPGPPDERAIELARVTRDGIRESALNLEKMWIVPGKIVRVVFI 138
Qy 138 DLICLDYDGNILDACTFALLAALKNVQLPEVTINEETALAEVNKKSYLNVRANPVATS 197
Db 139 DVHVLDDHGNLMDAIGIASTAALLNAKVPKVEVNEETGEVEI-LEEKEPLPVERIPDVT 197
Qy 198 FAVFDLTLIVDP-----TGEEGHPVHRNLNRSNGRQAVLSQARWEW 242
Db 198 FAKIGN-ILVVDPSLEBELVMDGRLTIVTDTDETH-----ISAVQKSEG 239
Qy 243 AAGAKLODCMSRAVTRHKEVSKLLDEVIQSMK 274
Db 240 GA-FKLEVMYAVETAFKAEIRKIVLEAIK 270

RESULT 4

F75181
polyribonucleotide nucleotidyltransferase related protein PAB0421 - Pyrococcus abyssi
C;Species: Pyrococcus abyssi
C;Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 09-Jul-2004
C;Accession: F75181
R;anonymous, Genoscope
submitted to the EMBL Data Library, July 1999
A;Description: Pyrococcus abyssi genome sequence: insights into archaeal chromosome stru
A;Reference number: A75001
A;Accession: F75181
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-274 <KAW>
A;Cross-references: UNIPROT:Q9V118; GB:AJ248284; GB:AL096836; NID:g5457730; PIDN:CAB4953
A;Experimental source: strain Orsay
C;Genetics:
A;Gene: PAB0421
C;Superfamily: conserved hypothetical protein MTH682

Query Match 25.0%; Score 359; DB 2; Length 274;
Best Local Similarity 34.5%; Pred. No. 1.1e-24;
Matches 95; Conservative 40; Mismatches 102; Indels 38; Gaps 6;

Qy 18 LKENCRRPDGRELGEFRRTTNIGSISTADGSAIVKLGNTTIVICGKAEFAAPPVADAPRG 77
Db 19 LKEGKRIDRGFDYRPIEIVGLIEKAGSALVKGSTQVLVGIKITLGEFFPTPNMG 78
Qy 78 YVNVNVDLPPLCSSRFRTPGPEEAQVTSQFIADVIENSHIIKKEDLCISPGKLAWLVC 137
Db 79 VMTTNVELVLASPTTFEPGPPDERAIELARVTRDGIRESALNLEKMWIVPGKIVRVVFI 138
Qy 138 DLICLDYDGNILDACTFALLAALKNVQLPEVTINEETALAEVNKKSYLNVRANPVATS 197
Db 139 DVHVLDDHGNLMDAIGIASTAALLNAKVPKVEVNEETGEVE-TLDETEPLPVEKIPVPT 197
Qy 198 FAVFDLTLIVDP-----TGEEGHPVHRNLNRSNGRQAVLSQARWEW 242
Db 198 FAKIGN-ILVVDPSLEBELVMDGRLTIVTDTDETH-----ISAVQKSEG 239
Qy 243 AAGAKLODCMSRAVTRHK---EVSKLDEVITQSMK 274
Db 240 GA-FKLEVMYAVETAFKAEIRKIVLEAVEKAK 273

RESULT 5

H69190
conserved hypothetical protein MTH682 - Methanobacterium thermoautotrophicum (strain Del
C;Species: Methanobacterium thermoautotrophicum
C;Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 09-Jul-2004
C;Accession: H69190
R;Smith, D.R.; Doucette-Stamm, L.A.; Deloughery, C.; Lee, H.; Dubois, J.; Aldredge, T.;
Qiu, D.; Spadafora, R.; Vicaire, R.; Wang, Y.; Wierzbowski, J.; Gibson, R.; Jiwani, N.
ki, S.; Church, G.M.; Daniels, C.J.; Mao, J.; Rice, P.; Noelling, J.; Reeve, J.N.
J. Bacteriol. 179, 7135-7155, 1997
A;Title: Complete genome sequence of Methanobacterium thermoautotrophicum Delta H: funct
A;Reference number: A69000; MUID:98037514; PMID:9371463
A;Accession: H69190

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 23, 2005, 12:23:53 ; Search time 99 seconds
(without alignments)
927.781 Million cell updates/sec

Title: US-10-736-892-12

Perfect score: 1437

Sequence: 1 MAAGFKTVEPLYRRFLKE.....TRHKEVSKLLDEVISQMKHK 276

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1424015 seqs, 332791073 residues

Total number of hits satisfying chosen parameters: 1424015

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications AA:*

- 1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
- 2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep.*
- 3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
- 4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
- 5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
- 6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep.*
- 7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
- 8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
- 9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep.*
- 10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep.*
- 11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep.*
- 12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep.*
- 13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep.*
- 14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep.*
- 15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep.*
- 16: /cgn2_6/ptodata/1/pubpaa/US10D_PUBCOMB.pep.*
- 17: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
- 18: /cgn2_6/ptodata/1/pubpaa/US11_NEW_PUB.pep.*
- 19: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
- 20: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	421	29.3	302	15	US-10-424-599-277230
2	410	28.5	307	15	US-10-424-599-277231
3	389	27.1	346	15	US-10-425-114-65625
4	389	27.1	347	15	US-10-425-114-63222
5	374	26.0	275	16	US-10-437-963-119834
6	371	25.8	334	15	US-10-425-114-47476
7	275	19.1	396	16	US-10-437-963-144343
8	267	18.6	430	15	US-10-425-114-64425
9	260.5	18.1	438	16	US-10-451-861-13
10	255.5	17.8	207	16	US-10-767-701-38982
11	236	16.4	108	15	US-10-108-260A-3952
12	233.5	16.2	287	15	US-10-310-154-625
13	203.5	14.2	323	15	US-10-424-599-254036

14	202	14.1	124	16	US-10-767-701-62295	Sequence 62295, A
15	199.5	13.9	286	15	US-10-424-599-254030	Sequence 254030, A
16	181	12.6	337	14	US-10-032-585-7279	Sequence 7279, Ap
17	176.5	12.3	154	15	US-10-424-599-261519	Sequence 261519, A
18	139.5	9.7	123	16	US-10-767-701-57278	Sequence 57278, A
19	134.5	9.4	359	15	US-10-282-122A-45707	Sequence 45707, A
20	131	9.1	350	14	US-10-032-585-7080	Sequence 7080, Ap
21	121	8.4	361	15	US-10-282-122A-57735	Sequence 57735, A
22	117	8.1	245	10	US-09-893-519A-48	Sequence 48, Appl
23	116	8.1	192	15	US-10-425-114-39165	Sequence 39165, A
24	116	8.1	278	15	US-10-425-114-62410	Sequence 62410, A
25	116	8.1	292	15	US-10-425-114-49949	Sequence 49949, A
26	115.5	8.0	168	16	US-10-437-963-129437	Sequence 129437, A
27	115	8.0	223	16	US-10-767-701-32772	Sequence 32772, A
28	115	8.0	260	15	US-10-425-114-61823	Sequence 61823, A
29	115	8.0	271	15	US-10-425-114-68490	Sequence 68490, A
30	113.5	7.9	245	15	US-10-258-662-1	Sequence 1, Appli
31	113.5	7.9	256	9	US-09-925-398-664	Sequence 664, App
32	113.5	7.9	256	14	US-10-102-806-664	Sequence 664, App
33	113	7.9	242	16	US-10-437-963-121395	Sequence 121395, A
34	112.5	7.8	241	15	US-10-424-599-260580	Sequence 260580, A
35	109	7.6	254	15	US-10-424-599-227970	Sequence 227970, A
36	108	7.5	191	15	US-10-264-237-2044	Sequence 2044, Ap
37	107	7.4	259	15	US-10-424-599-173123	Sequence 173123, A
38	104	7.2	362	15	US-10-282-122A-45571	Sequence 45571, A
39	101.5	7.1	118	16	US-10-437-963-168217	Sequence 168217, A
40	101	7.0	451	9	US-09-815-242-10721	Sequence 10721, A
41	100	7.0	451	15	US-10-282-122A-57142	Sequence 57142, A
42	99.5	6.9	705	10	US-09-907-907A-44	Sequence 44, Appl
43	99.5	6.9	705	10	US-09-907-907A-44	Sequence 44, Appl
44	98	6.8	274	15	US-10-425-114-50964	Sequence 50964, A
45	98	6.8	274	15	US-10-425-114-68491	Sequence 68491, A

ALIGNMENTS

RESULT 1

US-10-424-599-277230
; Sequence 277230, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 277230
; LENGTH: 302
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_92360C.1.pep
US-10-424-599-277230

Query Match	29.3%	Score	421	DB	15	Length	302
Best Local Similarity	32.2%	Pred. No.	3.1e-36				
Matches	92	Conservative	67	Mismatches	107	Indels	20
Gaps	4						
Qy	5	FKTVEPLYRRFLKENCPRDGRGELGEFRRTTNIGSISTADGSAVLKLGNTTTCGVKA	64				
Db	19	FRRLPLRYFERHIAESIRPDGRPLGKARETSIFLGAVASANGSALVKIGSTTILTAIKM	78				
Qy	65	EFAAPPVADPRGVVNVNVDLPPLCSSRFRTPGPERAQVTSQFIADVIENSHIHKEDL	124				
Db	79	EVMTPSLESDEGLAIDFHMPPICSPIVRGRPAEASPVVSKQSDTSSSKMIDLK	138				
Qy	125	CISPGKLAWLYCDLICLDYDGNILDGACTFALLAALKNVQLPEVTINEETALAEVN----	180				

```
Db 139 SLVSGKAAMAYLDIYCLDADGALFDALLSAVLSHLQIPVAVNDDGKIVLVEEDG 198
QY 181 -----LKKKSYLNVRANPVATSFVFDLTLIVDPTGEGHPVHRN-----LNRNGR 228
Db 199 KRQEQPVNKKRKLTLRSIPSLT-CILHKNYILADPTABEESIMETHVTIVLDTSDS--- 254
QY 229 GRQAVLSSQARWEAAGAKLQDCMSRAVTRHKEVSKLDEVIOQSMK 274
Db 255 GQLISLYKPGPVLAYTSAIQDCAALTRQVRKELKSFLDKANSAME 300
```

RESULT 2

```
US-10-424-599-277231
; Sequence 277231, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; FILE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 277231
; LENGTH: 307
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_92361C.1.pap
US-10-424-599-277231
```

```
Query Match 28.5%; Score 410; DB 15; Length 307;
Best Local Similarity 32.1%; Pred. No. 4, 8e-35;
Matches 93; Conservative 66; Mismatches 107; Indels 24; Gaps 5;

QY 5 FKTVEPLEYRRFLKENCRCRPGRELGEFRTTNTVIGSISTADGSALVKLGNTTVCVKA 64
Db 19 FRRLPFLYFFERHLAESIRPDGRLPGKARETSIFLGAVASANGSALVKIGSTMLTAIKM 78
QY 65 EFAAPPVADPRGYVNVNVDLPPLCSSRFRTPGPGEEAQVTSQFIADVIENSHIHKEDL 124
Db 79 EVMTPLSPPDGLAIDFHMPPICSPVIRPGRAEPASPVWSKQSDTSSRMIDLKEL 138
QY 125 CISPGLAWLYC-----DLICLDYDGNILDCTFALLAALKNVQLPEVTINEETAL 176
Db 139 SLVSGKAAMAYLNAAFSIMQDIYCLDADGALFDAALISAVAALSHLQIPAVAMNDDGKI 198
QY 177 AEVN-----LKKKSYLNVRANPVATSFVFDLTLIVDPTGEGHPVHRN-----LNR 224
Db 199 VLVSGEDGQKPVNKKRKLTLRSIPSLT-CILHKNYILADPTABEESIMETHVTIVLDT 257
QY 225 SNRGRAVLSSQARWEAAGAKLQDCMSRAVTRHKEVSKLDEVIOQSMK 274
Db 258 S---GQLISLYKPGGPALAYTSAIQDVGLTGQRVKELKSYLDKANSAME 304
```

RESULT 3

```
US-10-425-114-65625
; Sequence 65625, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE OF INVENTION: Plants and Uses Thereof for Plant Improvement
```

```
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 65625
; LENGTH: 346
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: 700242053_FLI.pap
US-10-425-114-65625
```

```
Query Match 27.1%; Score 389; DB 15; Length 346;
Best Local Similarity 30.6%; Pred. No. 1e-32;
Matches 86; Conservative 68; Mismatches 115; Indels 12; Gaps 2;

QY 5 FKTVEPLEYRRFLKENCRCRPGRELGEFRTTNTVIGSISTADGSALVKLGNTTVCVKA 64
Db 62 YRRLFPVAFLEHLCESVRIDARLRREARPTTVALGAVSSAHGSALARLGDATMLASVKL 121
QY 65 EFAAPPVADPRGYVNVNVDLPPLCSSRFRTPGPGEEAQVTSQFIADVIENSHIHKEDL 124
Db 122 EVMSPPAEHPDEGSVAVEFHMPPICSPVIRPGRPTDAPVISALEDVLTSSGMLNLKDL 181
QY 125 CISPGLAWLYC-----DLICLDYDGNILDCTFALLAALKNVQLPEVTINEETAL----- 176
Db 182 CLITGKASWLAYLDIYCLNADGSLFDAALISAVAAFTLHLEIPLVSGDGRFLTFTVGGNDG 241
QY 177 ---AEVNLKKSXYLNVRANPVATSFVFDLTLIVDPTGEGHPVHRNLSNRGROAV 233
Db 242 KNPVELYNREKRLTLGAILSLTCLHKKDLIL-ADPTSEESIEIETVTVVVDSDRLV 300
QY 234 LSSQARWEAAGAKLQDCMSRAVTRHKEVSKLDEVIOQSMK 274
Db 301 SLQKLGAVTCWATIKECISLAKERRRRSLREILDSIKAME 341
```

RESULT 4

```
US-10-425-114-63522
; Sequence 63522, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 63522
; LENGTH: 347
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: UC-ZMFLB73298B11_FLI.pap
US-10-425-114-63522
```

```
Query Match 27.1%; Score 389; DB 15; Length 347;
Best Local Similarity 30.6%; Pred. No. 1e-32;
Matches 86; Conservative 68; Mismatches 115; Indels 12; Gaps 2;

QY 5 FKTVEPLEYRRFLKENCRCRPGRELGEFRTTNTVIGSISTADGSALVKLGNTTVCVKA 64
Db 63 YRRLFPVAFLEHLCESVRIDARLRREARPTTVALGAVSSAHGSALARLGDATMLASVKL 122
QY 65 EFAAPPVADPRGYVNVNVDLPPLCSSRFRTPGPGEEAQVTSQFIADVIENSHIHKEDL 124
Db 123 EVMSPPAEHPDEGSVAVEFHMPPICSPVIRPGRPTDAPVISALEDVLTSSGMLNLKDL 182
```

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 23, 2005, 12:18:12 ; Search time 49 Seconds
(without alignments)
420.472 Million cell updates/sec

Title: US-10-736-892-12

Perfect score: 1437

Sequence: 1 MAAGFKTVBPLEYRFLKE.....TRHKEVSKLLDEVIOQMKHK 276

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 513545 seqs, 74649064 residues

Total number of hits satisfying chosen parameters: 513545

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued_Patents_AA.*

- 1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep.*
- 2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep.*
- 3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep.*
- 4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep.*
- 5: /cgn2_6/ptodata/1/iaa/PTUS_COMB.pep.*
- 6: /cgn2_6/ptodata/1/iaa/backfiles.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	257	17.9	291	4	US-09-976-594-326
2	187.5	13.0	224	4	US-09-270-767-32918
3	187.5	13.0	224	4	US-09-270-767-48135
4	183	12.7	349	4	US-09-248-796A-17015
5	136	9.5	392	4	US-09-248-796A-14627
6	121	8.4	471	4	US-09-107-532A-5705
7	100	7.0	455	4	US-09-134-000C-6074
8	96.5	6.7	252	4	US-09-252-991A-23776
9	95.5	6.6	259	4	US-09-489-039A-10604
10	94.5	6.6	257	4	US-09-543-681A-7481
11	91.5	6.4	721	4	US-09-949-016-6434
12	91.5	6.4	813	4	US-09-949-016-11077
13	91.5	6.4	1980	4	US-09-914-272A-3
14	91.5	6.4	1980	4	US-10-638-333-3
15	86.5	6.0	363	4	US-09-538-092-1072
16	86.5	6.0	1300	4	US-09-902-540-9932
17	85	5.9	1058	4	US-09-489-039A-11177
18	82	5.7	86	4	US-09-621-976-4688
19	82	5.7	244	4	US-09-328-352-5628
20	82	5.7	265	4	US-09-538-092-106
21	81.5	5.7	706	3	US-09-134-001C-4908
22	80.5	5.6	246	4	US-09-540-236-3210
23	80	5.6	304	4	US-09-902-540-15013
24	79	5.5	440	4	US-09-248-796A-25909
25	79	5.5	574	4	US-09-949-016-11325
26	79	5.5	737	4	US-09-583-110-2924
27	79	5.5	777	4	US-09-107-433-2700

28	79	5.5	1052	4	US-09-252-991A-30591	Sequence 30591, A
29	78.5	5.5	699	4	US-09-949-016-6073	Sequence 6073, Ap
30	78.5	5.5	720	4	US-09-949-016-9819	Sequence 9819, Ap
31	78.5	5.5	750	4	US-09-107-532A-5868	Sequence 5868, Ap
32	78.5	5.5	1023	4	US-09-514-907A-6	Sequence 6, Appli
33	78.5	5.5	1023	4	US-09-896-594-6	Sequence 6, Appli
34	78.5	5.5	1858	4	US-09-902-540-12643	Sequence 12643, A
35	78	5.4	60	4	US-09-513-999C-5994	Sequence 5994, Ap
36	78	5.4	552	2	US-08-317-401E-4	Sequence 4, Appli
37	77.5	5.4	448	4	US-09-409-096-4	Sequence 4, Appli
38	77.5	5.4	1088	4	US-09-233-857-4	Sequence 4, Appli
39	77.5	5.4	1212	3	US-09-090-535-1	Sequence 1, Appli
40	77.5	5.4	1212	3	US-09-090-535-2	Sequence 2, Appli
41	77.5	5.4	1212	3	US-09-090-535-3	Sequence 3, Appli
42	77.5	5.4	1212	3	US-09-090-535-4	Sequence 4, Appli
43	77.5	5.4	1633	4	US-09-902-540-12892	Sequence 12892, A
44	77.5	5.4	1971	4	US-09-914-272A-1	Sequence 1, Appli
45	77.5	5.4	1971	4	US-10-638-333-1	Sequence 1, Appli

ALIGNMENTS

RESULT 1
US-09-976-594-326
; Sequence 326, Application US/09976594
; Patent No. 6673549
; GENERAL INFORMATION:
; APPLICANT: Furness, Michael
; APPLICANT: Buchbinder, Jenny
; TITLE OF INVENTION: GENES EXPRESSED IN C3A LIVER CELL CULTURES TREATED WITH STEROIDS
; FILE REFERENCE: PA-0041 US
; CURRENT APPLICATION NUMBER: US/09/976,594
; PRIOR FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: 60/240,409
; PRIOR FILING DATE: 2000-10-12
; NUMBER OF SEQ ID NOS: 1143
; SOFTWARE: PERL Program
; SEQ ID NO 326
; LENGTH: 291
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. 6673549 2551987CDI
US-09-976-594-326

Query Match	17.9%	Score	257	DB	4	Length	291
Best Local Similarity	27.1%	Pred. No.	3.2e-22				
Matches	76	Conservative	47	Mismatches	115	Indels	42
Gaps	7						
Qy	18	LKENCPRDRELGBFRITTTVNI	IGSISTADGSALVKLGNTTVICG	KAFAAAPPVDPADPRG	77		
Db	18	VQEDLRVDGRCEDYRCVETDVT	VVNTSGARVLGHTDLVGVKAEMGT	PKLEKPNEG	77		
Qy	78	YVWPNVDLPCLCSRFRTPGPEE	-AQVTSQFIADVIENSHIIKKBDI	CISPGKLAWLY	136		
Db	78	YLEPFVDCSASATPEFE-GRGD	DLGTETANTLVRIENKSSVDLKT	LICISPREHCWLY	136		
Qy	137	CDLICLDYDGNLDACFTFALLA	LKXVOLPEV-TINEETALAEVNLK	KSY----	LNURA	191	
Db	137	VDVLLLECGGNLFDAISIAVKA	ALFNTRIPRVRLDEEGSKDIELSD	DDPYDCIRLSVEN	196		
Qy	192	NPVATSPA-----VFDDTL	-----LIVDPTGEGHPVHRLN	RSNRCGRQAVLSS	236		
Db	197	VPCIVTLCKTYRHVVDATIQE	ACSLASLVSVTSKGVVTCMRKV	GK-----	244		
Qy	237	QARWEWAAGAKLQDCMSRAV	TRHKEVSKLLDEVIOQMKHK	276			
Db	245	-----GSLDPESIFEMMETG	KRVGKVLHASLQSVLHK	276			

RESULT 2

[illegible]

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 23, 2005, 12:09:32 ; Search time 67 seconds
(without alignments)
1593.222 Million cell updates/sec

Title: US-10-736-892-12

Perfect score: 1437

Sequence: 1 MAAGFKTVEPLEYRRFLKE.....TRHKEVSKLDEVIQSMKHK 276

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2105692 seqs, 386760381 residues

Total number of hits satisfying chosen parameters: 2105692

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A Geneseq_16Dec04.*

- 1: Geneseqp1980s.*
- 2: Geneseqp1990s.*
- 3: Geneseqp2000s.*
- 4: Geneseqp2001s.*
- 5: Geneseqp2002s.*
- 6: Geneseqp2003as.*
- 7: Geneseqp2003bs.*
- 8: Geneseqp2004s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB ID	Description
1	1184	82.4	276	5	ABB97415
2	1184	82.4	276	8	ABM81511
3	1184	82.4	276	8	ADP55135
4	806	56.1	171	4	AB63406
5	806	56.1	171	4	AB63276
6	773	53.8	157	4	AAB63405
7	431	30.0	289	3	AA293331
8	431	30.0	302	3	AA293330
9	431	30.0	352	3	AA293329
10	362	25.2	112	4	AB63407
11	360	25.1	113	8	ADR08498
12	354.5	24.7	272	8	ADN47715
13	316.5	22.0	267	7	ADM25774
14	271.5	18.9	305	6	ABR52963
15	271.5	18.9	305	7	ADK62640
16	264	18.4	300	3	AGS51508
17	264	18.4	300	3	AG223307
18	264	18.4	307	3	AGS51507
19	264	18.4	307	3	AG223306
20	260.5	18.1	438	5	ABB83339
21	257	17.9	291	7	ADC31596
22	257	17.9	291	8	ADL12597
23	257	17.9	291	8	ABM80445
24	257	17.9	291	8	ADP55026
25	257	17.9	299	7	ADC33202

ALIGNMENTS

RESULT 1

ABB97415

ID ABB97415 standard; protein; 276 AA.

XX ABB97415;

XX 27-JUN-2002 (first entry)

XX Novel human protein SEQ ID NO: 683.

XX Human; antianaemic; vulnery; antinflammatory; immunomodulator;

KW antiinfertility; cerebroprotective; cytostatic; rheumatic; gene therapy;
KW neuroprotective; antiparkinsonian; protein therapy; EST;
KW expressed sequence tag.

XX Homo sapiens.

XX WO200222660-A2.

XX 21-MAR-2002.

XX 10-SEP-2001; 2001WO-US026015.

XX 11-SEP-2000; 2000US-00659671.

XX (HYSE-) HYSEQ INC.

PI Tang YT, Liu C, Zhou P, Asundi V, Zhang J, Zhao QA, Ren F;

PI Xue AJ, Yang Y, Wehrman T, Drmanac RT;

XX WPI; 2002-292408/33.

XX N-PSDB; ABN32601.

PT An isolated polynucleotide for treating diseases associated with its

PT encoded polypeptide such as cancer and multiple sclerosis.

XX Claim 20; SEQ ID NO 683; 509pp; English.

CC The present invention provides the protein and coding sequences of 444
CC novel human proteins. These were isolated from expressed sequences tags
CC (ESTs). They can be used to stimulate cell growth, to regulate
CC haematopoiesis e.g. to treat aplastic anaemia, to help tissue regrowth
CC e.g. in burn treatment, to regulate the immune system e.g. to treat
CC multiple sclerosis, to regulate activin or inhibin e.g. to treat
CC infertility, to regulate haemostasis or thrombolysis e.g. to treat stroke
CC and cancer, to screen for drugs, to treat inflammatory conditions e.g.
CC rheumatoid arthritis, and to treat nervous system disorders e.g.
CC Parkinson's disease. The present sequence is a protein of the invention

```

XX SQ Sequence 276 AA;
Query Match 82.4%; Score 1184; DB 5; Length 276;
Best Local Similarity 81.3%; Pred. No. 4.6e-127;
Matches 235; Conservative 9; Mismatches 19; Indels 26; Gaps 2;

QY 1 MAAGFKTVEPLEYRRFLKENCPRDGRGELGFEFRITTTVNIIGSISTADGSALVKLGNTTVC 60
DB 1 MAAGFKTVEPLEYRRFLKENCPRDGRGELGFEFRITTTVNIIGSISTADGSALVKLGNTTVC 60

QY 61 GVKAEFAAPPVADPRGYVVPNDLPPLCSSRFRTPGPEEAQVTSOFTIADVIENSHIK 120
DB 61 GVKAEFAAPPVADPRGYVVPNDLPPLCSSRFRTPGPEEAQVTSOFTIADVIENSHIK 120

QY 121 KEDLCISPGKLAWLYCDLICLDYDGNILDCTFALLAALKNVQLPEVTINEETALAEVN 180
DB 121 KEDLCISPGKLAWLYCDLICLDYDGNILDCTFALLAALKNVQLPEVTINEETALAEVN 180

QY 181 LKKKSYLVNRPVATSFVFDLTLIVDPTGEEGHPVHRN-----LNRSG 227
DB 181 LKKKSYLVNRPVATSFVFDLTLIVDPTGEEGHPVHRN-----LNRSG 227

QY 228 RGRQAVLSSQARWEAAGAKLQDCMSRAVTRHKEVSKLLDEVIOQMKHK 276
DB 241 SG-----LTGAKLQDCMSRAVTRHKEVSKLLDEVIOQMKHK 276

RESULT 3
ADP55135
ID ADP55135 standard; protein; 276 AA.
XX AC ADP55135;
XX DT 18-NOV-2004 (first entry)
XX DE Human PRO protein sequence SEQ ID NO:1111.
XX KW human; PRO; immune related disease; inflammatory immune response;
KW immune response stimulation; antiallergic; antianaemic; antiarthritic;
KW antiaesthetic; antidiabetic; antineoplastic; antipsoriatic;
KW antirheumatic; antithyroid; CNS; dermatological; gastrointestinal;
KW haemostatic; hepatotropic; immunostimulant; immunosuppressive; muscular;
KW nephrotropic; neuroprotective; osteopathic; respiratory; vasotropic;
KW virucide; gene therapy.
XX OS Homo sapiens.
XX PN WO2004039956-A2.
XX PD 13-MAY-2004.
XX PF 28-OCT-2003; 2003WO-US034381.
XX PR 29-OCT-2002; 2002US-0422472P.
XX CC overexpressed in cancer tissues compared to normal tissues, and may thus

```

```

XX SQ Sequence 276 AA;
Query Match 82.4%; Score 1184; DB 5; Length 276;
Best Local Similarity 81.3%; Pred. No. 4.6e-127;
Matches 235; Conservative 9; Mismatches 19; Indels 26; Gaps 2;

QY 1 MAAGFKTVEPLEYRRFLKENCPRDGRGELGFEFRITTTVNIIGSISTADGSALVKLGNTTVC 60
DB 1 MAAGFKTVEPLEYRRFLKENCPRDGRGELGFEFRITTTVNIIGSISTADGSALVKLGNTTVC 60

QY 61 GVKAEFAAPPVADPRGYVVPNDLPPLCSSRFRTPGPEEAQVTSOFTIADVIENSHIK 120
DB 61 GVKAEFAAPPVADPRGYVVPNDLPPLCSSRFRTPGPEEAQVTSOFTIADVIENSHIK 120

QY 121 KEDLCISPGKLAWLYCDLICLDYDGNILDCTFALLAALKNVQLPEVTINEETALAEVN 180
DB 121 KEDLCISPGKLAWLYCDLICLDYDGNILDCTFALLAALKNVQLPEVTINEETALAEVN 180

QY 181 LKKKSYLVNRPVATSFVFDLTLIVDPTGEEGHPVHRN-----LNRSG 227
DB 181 LKKKSYLVNRPVATSFVFDLTLIVDPTGEEGHPVHRN-----LNRSG 227

QY 228 RGRQAVLSSQARWEAAGAKLQDCMSRAVTRHKEVSKLLDEVIOQMKHK 276
DB 241 SG-----LTGAKLQDCMSRAVTRHKEVSKLLDEVIOQMKHK 276

RESULT 2
ABM81511
ID ABM81511 standard; protein; 276 AA.
XX AC ABM81511;
XX DT 18-NOV-2004 (first entry)
XX DE Tumour-associated antigenic target (TAT) polypeptide PRO82291, SEQ:3894.
XX KW Tumour-associated antigenic target; TAT; human; overexpression; cancer;
KW tumour; diagnosis; cell proliferative disorder; breast cancer;
KW colorectal cancer; lung cancer; ovarian cancer; liver cancer;
KW central nervous system cancer; bladder cancer; pancreatic cancer;
KW cervical cancer; melanoma; leukaemia; hybridisation probe;
KW chromosome identification; chromosome mapping; gene mapping;
KW gene therapy; cytostatic.
XX OS Homo sapiens.
XX PN WO2004030615-A2.
XX PD 15-APR-2004.
XX PF 29-SEP-2003; 2003WO-US028547.
XX PR 02-OCT-2002; 2002US-0414971P.
XX PA (GETH ) GENENTECH INC.
XX PI Wu TD, Zhang Z, Zhou Y;
XX WP1; 2004-347921/32.
XX DR N-PSDB; ACN39626.
XX PT New tumor-associated antigenic target polypeptides and nucleic acids,
PT useful in preparing a medicament for treating or detecting a
PT proliferative disorder, e.g. breast, lung, colorectal, ovarian or
PT prostate cancer or tumor.
XX PS Claim 12; SEQ ID NO 3894; 7273pp; English.
XX CC The invention relates to human tumour-associated antigenic target (TAT)
XX polypeptides, and their related nucleic acids. The TAT polypeptides are
XX overexpressed in cancer tissues compared to normal tissues, and may thus

```